ESSAYS

n. 12/2002

Conference on Agricultural and Environmental Statistical Applications in Rome

Rome, 4-8 June 2001

ISTAT

With the collaboration of:

Ministero delle Politiche Agricole e Forestali

eurostat

NASS

OECD OCDE

UN/ECE

Editing by Federica Piersimoni (Istat)

Editing review: Cecilia Manzi (Istat) and Marcel Van den Broecke (ISI)

Conference on Agricultural and Environmental Statistical Applications in Rome

Essays n. 12 - 2002

National Statistical Institute of Italy
Via Cesare Balbo, 16 - Roma

Editorial co-ordination:
Piero Crivelli
Servizio Produzione editoriale

Layout: Vittorio Cioncoloni, Piero Crivelli

Cover: Vittorio Cioncoloni

CD ROM: Vittorio Cioncoloni

Printing:
Poligrafica Ruggiero s.r.l.
Zona industriale pianodardine - Avellino
Copies: 2300

Table of Contents

Volume I - 5 June 2001

FISHERY STATISTICS: CURRENT AND FUTURE CHALLENGES

Volume II - 6 June 2001

Plenary Session II

QUALITY STANDARDS FOR AGRICULTURAL STATISTICS

Invited Paper Sessions

AGRICULTURAL CENSUS AND SURVEY ISSUES

TIMELINESS/ACCURACY/COMPLETENESS OF AGRICULTURAL STATISTICS

AGRICULTURAL STATISTICS SYSTEMS

CLASSIFICATION ISSUES FOR AGRICULTURE STATISTICS SYSTEM

AGRICULTURAL SURVEY METHODS IN TRANSITION AND DEVELOPING COUNTRIES

DATA SOURCES AND DESIGN ISSUES FOR AN AGRICULTURE STATISTICS SYSTEM

DATA INTEGRATION ISSUES

Volume III - 7 June 2001

Plenary Session III

TECHNOLOGICAL IMPACT ON DATA MANAGEMENT

Invited Paper Sessions

ADVANCED SURVEY DESIGN AND ESTIMATION ISSUES

NEW TECHNOLOGIES AND STRATEGIES FOR AGRICULTURAL DATA COLLECTION AND FOR USE IN ENVIRONMENT MONITORING

STATISTICAL DATA EDITING: METHODS AND SOFTWARE

AGRICULTURAL MODELLING

Volume IV

Contributed Paper Sessions

5 June 2001

AGRICULTURAL CENSUSES

AGRICULTURAL STATISTICAL SYSTEMS

6 JUNE 2001

THE AGRICULTURE-ENVIRONMENT RELATIONSHIP

DATA PROCESSING

GIS AND OTHER TECHNOLOGICAL ASPECTS

Scientific Committee:

M. Anastasiadis, Eurostat
R. Benedetti, Istat
D. Berze, ISI
G. Calò, Eurostat
J Gallego, JRC
E. Giovannini, OECD
Á. González Villalobos, FAO
L. Kabat, FAO
J. Karlsson, UN/ECE
A. Lindner, OECD
A. Mancini, Istat
R. Muthmann, Eurostat
P. Narain, FAO
F. A. Vogel, NASS/USDA

Organizer Committee:

Andrea Mancini
Roberto Benedetti
Giovanni Cicchetti
Maurizio Esposito
Cristina Forlacroix
Maria Frustaci
Laura Machetti
Cinzia Pellicanò
Leopoldo Pennese
Federica Piersimoni
Giuseppe Serino
Giuseppe Stassi

ESSAYS

n. 12/2002

Conference on Agricultural and Environmental Statistical Applications in Rome

Rome, 4-8 June 2001

Volume I

ISTAT

FOREWORD

In June 2001, the international “Conference on Agricultural and Environmental Statistical Applications” (CAESAR) was held in Rome, Italy. The conference, modelled after the similar meeting organised in Washington D.C. in 1998, brought together 400 statisticians and policy experts from 53 countries around the world and from all the main international organisations active in the field of agriculture.

The Italian Institute of Statistics (ISTAT) and the Italian Ministry of Agricultural and Forestry Policies (MIPAF) hosted the event, which was co-organised by the Statistical Office of the European Communities (Eurostat), the International Statistical Institute (ISI), the Economic Commission for Europe of the United Nations (UN/ECE), the National Agricultural Statistics Service of the United States Department of Agriculture (NASS), the Organisation for Economic Co-operation and Development (OECD) and the Food and Agriculture Organisation of the United Nations (FAO). All these institutions co-operated very closely to achieve a successful event, and contributed to the activities of the Scientific Committee. A local Organiser Committee dealt with all the logistic and organisational issues. On behalf of all participants, I would like to express my sincere gratitude to both the Scientific and the Organiser Committees for their excellent work.

The Conference program was divided into five plenary sessions, 17 invited paper sessions and 7 contributed paper sessions. Each day was devoted to analysing a particular perspective of the development of agricultural statistics: the demand for agricultural statistics in the new millennium (both for developed and developing countries), methodological issues and new technologies for collecting, processing and disseminating agricultural statistics. Interactions between agriculture, the environment and the social aspects were discussed as well. In addition, two satellite meetings were organised to discuss in detail the situation of agricultural statistics in transition countries and the FAOSTAT program. In total, about 120 papers were presented and discussed during the Conference.

This volume contains the proceedings of the Conference: in particular, five opening speeches, sixty invited papers and forty-six contributed papers are included, as well as six papers presented in the two satellite meetings.

In conclusion, the CAESAR Conference was an important event for helping national and international organisations to exchange innovative ideas, identifying current and future users’ needs, and comparing best practices related to the adoption of new technologies for agriculture statistics. It confirmed the need for having regular opportunities for bringing together, at a high level and in an international environment, users and producers of agriculture statistics: all international organisations involved in the preparation of the Conference committed themselves to organise a “Third International Conference on Agricultural Statistics” in 2004.

Enrico Giovannini
Chief Statistician of the OECD and
Chairman of the CAESAR Scientific Committee

5 June 2001

Plenary Session I

The Agricultural Situation in the New Millennium

Chair: L. Biggeri

Opening Ceremony

L. Biggeri
President of Istat - Via C. Balbo, 16 - Roma

As President of ISTAT I am particularly pleased to offer all you a warm welcome and a sincere thanks to the delegates and observers coming from all over the world to attend the CAESAR Conference on Agricultural and Environmental Statistics.
Moreover I would like to thank all the Agencies and Institutions involved in the organization of this event: EUROSTAT, the National Agricultural Statistics Services for United States Department of Agriculture (NASS), The Organization for Economic Co-operation and Development (OECD), The International Statistical Institute (ISI), The United Nations and the Food and Agriculture Organization (FAO).
Today we start one of the most important meetings in terms of number and quality of participants devoted to agriculture statistics. It represents a follow up to the successful first conference held in Washington DC in March 18-20, 1998. During this week we expect more than 300 participants representing about 70 countries attending the scientific sessions and the satellite meetings.
The objective of this conference is to stimulate both the development of an information network of individuals, institutions involved in the use and production of agricultural statistics and the analysis of emerging issues in agriculture, especially those concerning environment and rural development.
At the turning point of a new millennium, this conference represents a unique opportunity for statisticians, agricultural experts, users and other researchers to meet, discuss the status of the statistical research in agriculture and focus both on future needs for agricultural statistics and on the requirements to meet those needs for statistical technological development.
The importance of agriculture in our society is relevant and the role of the statisticians has to be adequate to it.
The health and wealth of a nation and its potential to develop and grow depend upon its ability to feed its people.
Agriculture, as a primary activity, is directly connected to food availability. In a world of 6 billion people, where the population is expected to double in 50 years, the problem of alimentation is, already, one of the central and strategic themes of our society. But at the same time, the possible answers of agriculture to this question such as the productivity increases or the use of new areas for cultivation and livestock, have a cost in terms of sustainable development. Environmental and social impact involved in the transformation of agriculture have to be considered and studied. More and more, the world faces "crises" related to the quality of food products and the risks linked to rapid progress in biotechnology. "Mad cow" disease, dioxins in the food chain, and disagreements over trade in genetically modified organisms are serious causes of concern for governments and public debate.
According to these considerations, the priority is monitoring the evolution of the agricultural activities in order to control the global development process of planet Earth.
The situation is very different among countries and this gives rise to distinct statistical approaches in the primary sector. In the developing countries the agricultural activity is more oriented to rural development and to ensure the basic alimentation for the

population than in the developed countries where the agriculture involves, also, many other aspects related to food quality, environmental impact, market destinations and their connected services.

In this picture, the role of the statistician is very crucial and complex one. Accurate and timely statistics about the source and availability of basic agricultural supplies are essential. In recent years, world economies have become increasingly interdependent, and the globalisation of economic activities has expanded the need for statistics about agriculture. More specifically official statistics have different tasks:

- ✓ to provide a detailed and comparable picture of the structure of the sector among geographical areas through the censuses;
- ✓ to offer timely and reliable information on the markets to follow its evolution in terms of productions, values and commercial exchanges;
- ✓ to satisfy the growing statistical demand on all the important aspects related to the rural development, the occupation, the environment and the food quality.

As you can see in the scientific program, the conference agenda during this week is very full and divided into plenary, invited and contributed parallel sessions in which more than 100 papers cover, in my opinion, the most relevant topics in agriculture statistics.

Today the discussion is focused on the description of the current situation and the perspectives of agriculture; in the afternoon parallel sessions its role in the world economies and its social changes besides some statistical applications on the environmental impact and on fishery will be analyzed.

Tomorrow the sessions will be mainly devoted to the techniques, the instruments and the modalities used by official statistics to monitor and analyse agricultural activities as well as to answer to each information need.

On Thursday the conference will give an overview of the applied research in the field of technological and methodological solutions to the problems emerged from agricultural statistics.

1. An overview of 5th June sessions: the agro-environmental situation and the user needs

The economic history of the second half of the 20th Century clearly presents the changes taking place in the agricultural sector and the speed of change we should expect for this millennium. Production and demand of agricultural commodities exponentially increased after the world war II, following the expansion of global market and an intense exploitation of natural resources. Wheat, fish, meat and wood, for instance, doubled in just fifty years (data from the *Meat World Congress* 1999). Such a relevant jump in production and exploitation of resources generated complex transformations in farms organization and management, production techniques, tastes of consumers and, in general, way of life around the world.

The consequence of such a wide transformation is that the nature of agriculture itself is changing over time: agricultural producers display a wide heterogeneity in different countries and, moreover, within the same country too; environment and natural resources play an increasing role in the sector; the production of nutritious and safe food is a question involving developing and, more and more, developed economies.

The changes of the primary production process have, moreover, heavily conditioned the territorial order of most regions, upsetting the previous natural balance. At the beginning of the new millennium the Earth's conditions are not healthy. Serious problems affect developed, developing and regressing areas: pollution, urban concentration, desertification, lack of natural resources and climate changes, are only the most well-known ones.

The global dimension of such important items suggest that it is no longer possible to approach the agricultural sector without considering its natural foundations.

Agriculture is an important factor of social and economic growth in all regions, even though with different meanings in the North and in the South of the world. In highly productive areas, the major issues are water pollution from nutrients, herbicides, and fungicides; soil erosion; loss of botanical diversity; land degradation caused by abandonment and new and unknown risk for the human health. In low productivity areas different questions involve economic and natural factors, such as persistent poverty, hunger, malnutrition and food deficiencies, pollution of drinkable water, reduction of soil fertility and forest resources.

Special mention should be made of the previous planned economies of central Europe and ex Soviet Union, now rapidly moving toward a free market economy integrated with the world exchange system. Several countries in transition economy are facing problems similar to developed as well as poor countries. The structure of farms is very different in these economies: there are still large state farms next to small family ones. As a consequence, it is estimated that an important part of the agricultural production on the private farms is created in an environment-friendly way. Moreover, it will be extremely relevant to have firm control over the present state and the evolution of environmental conditions in such areas of the world after a long period of planned growth of production and exploitation of soil. External conditioning coming from industry and energy exploitation generated shocks to agriculture and environment without clear measures and understanding.

The relevance of these topics is so urgent that a special satellite meeting has been devoted to them during the opening day of this conference.

In all land exploited by humans, excluding urban or wilderness area, agriculture impacts directly or indirectly on environmental quality, becoming at the same time the degradation or the conservation factor. Maintaining social, economic and strategic value of agricultural sector in developed countries and assuring food security in developing regions requires the application of new sustainable development ideas: a sort of economic growth of the present generation without loss and degradation of natural resources to be left to future generations.

A good practice in agriculture, ensuring acceptable income for farm households and improving production of food and primary goods, is an actual main goal that requires pollution and waste control, a correct management of water and nutrients cycles, the maintenance of forest resources, botanical diversity and rural landscape. In such a way a good quality and a durable production in the poorest countries may be ensured together with a social and natural value of rural land in the richest areas.

Such an approach requires a statistical framework able to surmount the traditional dualism between agricultural production data system and statistical reports on environmental quality.

In this framework much work has been done and much more remains to be done for implementing the concept of quality and its main components to meet the emerging more sophisticated users' needs. Features like availability, clarity, confidentiality and security in addition to requirements as for the traditional components of quality in statistical data (relevance, accuracy, timeliness, accessibility, coherence, comparability, completeness) have been recently introduced also in the production process of agricultural statistics.

Many international agencies have strongly contributed to this process, the Food and Agricultural Organization (FAO), the Statistical Office of the European Communities (EUROSTAT), the United Nations/Economic Commission for Europe (UN/ECE), the National Agricultural Statistics Service (NASS) and the Organization for Economic Cooperation and Development (OECD) among others. All these organizations have promoted initiatives for monitoring users' informative demand and producing harmonized statistics to meet the new challenges in agricultural field.

As the agricultural and the food supply sectors evolve under long-term trends, the framework of agricultural statistics should provide the policy makers with necessary information to understand and to quantify the impact of economic policies in this evolving context. This issue is increasingly becoming relevant as there is a reconsideration of agriculture's contribution to growth and development: agriculture could be an engine for the sustainable economic development in the next millennium.

The scientific community has to bring about this structural evolution in the agricultural world and re-modulate definitions and instruments as required. A new definition of agriculture should be able to account for new characteristics and issues involved. The basic units of economic observation (farm, farm household, rural areas, environment, etc.) should be reconsidered to avoid obsolescence. There is a need for more information on non-farm businesses in rural area, regional statistics, agri-industry and agri-environmental statistics, input-output flows between sectors, supply chains and so on.

The underlined issues imply, in the future, the collection of better data; a greater integration of existing data on several subjects and a quick feedback for policy interventions. Of course, a deeper understanding of the processes at work requires better land and geographical information to be integrated with quantitative and qualitative data on the economy as well as on the natural environment. There must be, then, a continued discussion between users of statistics and providers of statistics, to realize a real link on needs of information and official production of data. In other words, the statistical system has to become more reactive to future needs.

2. An overview of 6th June sessions: main sources of agricultural statistics

In the past, most of the statistical information on the agricultural field came from censuses. Currently a renewing process has been carried out involving not only contents but also methodological aspects of the statistical production process so that nowadays different sources provide the core of agricultural statistical information in the various countries. These sources basically include censuses, sample surveys and administrative data.

An agricultural census has the main aim of providing a complete enumeration of the farm holdings in a country. It allows to provide a representation of the structure of farming and agricultural activity of a nation at a particular time. Many arguments still support the crucial role of a census in the agricultural statistical system of a country. They derive from the need of:

- ✓ establishing a list of all known agricultural holdings for sample surveying;

- ✓ establishing benchmark reference levels for the annual survey program;
- ✓ providing details for small geographical areas;
- ✓ providing information on commodities where production is limited.

Nonetheless a census has also limitations. They mainly consist of high costs in terms of financial and human resources that derive from conducting a large investigation in a relatively short period of time and from dealing with respondents with different levels of education.

In its guidelines for the World Census of Agriculture 2000, the FAO suggests to conduct a census at least every ten years on a complete coverage basis. The international agency also underlines the importance of enlarging information about emerging areas of interest such as employment, role of women in agriculture and environmental and sustainable development issues. However, individual countries adopted these broad rules according to their own needs and resources. On the one hand, developed countries have a consolidated tradition in conducting censuses dating back to the first Roman demographic enumeration. Nevertheless, some methodological issues still remain to be solved especially for coverage, classification and duplication. Currently these concerns represent the main areas of research in the census field. On the other hand, other priorities have to be faced by developing countries where budget limitations and agricultural practice, giving rise to very many small holders, have oriented the collection scheme towards a "large sample" as a practical strategy for covering the sector. The advantage of this practice is the opportunity of measuring additional information beyond the structural data such as production of crops and demographic characteristics of the agricultural population. As an example, the sampling scheme will be adopted in the next agricultural census of Indonesia and Nepal, while Pakistan, Philippines, Thailand and Sri Lanka, among others, planned to carry out a census based on a selected part of the agricultural holdings population. On the other side, China is one of the developing countries that conducted a complete enumeration census and plans a ten year periodical repetition of the enquiry.

Even though censuses play a central role in the Agricultural Statististical System, sample surveys strongly contributed, especially in recent years, to provide a more detailed overview on both structural and conjunctural aspects of farming activity with special reference to livestock, production means, economic accounts, price in agriculture. It is well-known that the main advantage of a sampling scheme is allowing an improvement of final estimates efficiency, simplifying data collection and facilitating data editing. In addition it allows to control both response and non-response errors by sampling design.

Sample surveys on many specific sub-sectors of agriculture have been carried out especially in developed countries. Eurostat promoted the development of harmonized questionnaires within EU member states with the main aim of providing support to the European Common Agriculture Policies. The investigations concern both traditional areas in agricultural production and emerging sub-sectors. A leading survey at Eurostat is the farm structure survey that provides on a biannual basis general information on the farms and data on cultivation, livestock, mechanical equipment, work used in the farm and external relations. To extend the scope of the survey beyond the usual agricultural domain and to include also aspects linked to environment, multi-functionality, landscape and sustainable development, specific programs have also been launched. LUCAS and TAPAS represent some example of these projects.

Finally administrative data coming from various institutional organizations are beginning to be integrated within the Agricultural Statistical System. Although statistical surveys and censuses remain the principal sources of information, there is now the opportunity of making good use of administrative data. The challenge is to maximize the use of different sources minimizing duplication and errors while at the same time avoiding contemporary, statistical annoyance. In the agricultural field administrative data mainly come from farm registers set up for monitoring fund applications and for many other organizative purposes.

The main advantages in using the administrative sources rely on the low additional cost for usage and on the opportunity of reducing the burden for agricultural holdings of providing data with a strong impact on the possibility of improving statistical services and quality of released statistics.

The disadvantages, mainly related to completeness, comparability and timeliness, derive from the scope the data are collected for, usually extending beyond statistics. Administrative data are indeed collected with the aim of control, while the statistical data target is monitoring and forecasting results in agricultural field.

Despite these disadvantages, the use of administrative data is strongly recommended since they don't need to be collected ad hoc since they are already available. Even though it seems to be an additional cost switching administrative data to statistical purposes, in a long period of time, it could be a strategic choice. The most critical point is merging the two systems: the statistical and the administrative one. Definitions, classification and contents, indeed, are rarely standardized, especially in the agriculture field, and need to be harmonized to get reliable figures. Some countries have started to gain experience in the use of administrative data. This is the case in the Integrated Administration and Control System (IACS) of Germany with data collected in connection with the Agricultural Reform and the Community Vineyard Register and in the Italian Statistical Archive of Active Farms (ASIA) which includes indicative data and basic information on industrial farms and agricultural farms in particular (ASIA-Agriculture). In addition the Farm Accountancy Data Networks (FADN) that provide information on budgeting and new needs in agricultural field, has recently been used for statistical purposes in various Countries.

The analysis of the world agriculture statistics current scenario gives rise to some considerations. Due to the emergence of new topics and the increased association of non-farm activities with the more traditional agricultural occupations, integration of different data sources has become a crucial requirement to produce complete and high quality statistics in the field of agriculture. The development of Agricultural Statistical Systems, involving more than isolated information on specific sub-sectors, is required to provide a general overview of farming activities and a better understanding of the sector and of its emerging issues. Also the adoption of a territorial approach rather than a farm based approach could contribute to focus on the integration of information on agricultural structure with a broader picture of economy. For this scope, the application of remote sensing and Geographical Information Systems as well as the role of Information Technology has become clearer in recent years. Despite various constraints, many countries, even developing ones, have started to implement these technological developments.

3. An overview of 7th June sessions: technological and methodological solutions

Stimulated by an increasing demand for information, national statistical institutes felt the need to improve their production process, in terms of external quality or relevance, that is the suitability of the data to meet users' requirement for content, informative detail, timeliness, easy of use, etc. On the other hand, official statisticians have to take care of three important aspects: the first one is the necessity to reduce the statistical burden on respondents, the second one is to provide reliable small area statistics to satisfy both public and private sectors' demand, and the last one is to correlate the survey methodology to the ongoing innovation in the field of the computer technologies.

All over the world, the solution to these problems involves more than one phase of the production process, namely: the dissemination of statistical information, the survey design methodology and the choice of estimators, the data collection and the methodology for error detection and correction.

Dissemination can be seen as the end point of the production process but also as the specific mission of official statistics in meeting users' needs for information. Fortunately, the technological evolution makes information plentiful and accessible, in the meantime orienting users' demand to quality and services rather than to quantity and products. Thus, the trends in statistical dissemination are towards widening and improving data supply, releasing them online through the internet, integrating information (contents, editorial standards, meta-information), lowering the barriers to access (both in terms of easy use and of costs).

In this context, statistical web-sites have a growing importance: they make it possible to disseminate huge volumes of data at low or at no cost, with a consistent interface; to segment the available information according to the level and the interests of different categories of users; to provide tools to navigate and extract custom-made information for the user's specific needs on a timely basis; and – last but not least – to build up statistical culture and awareness, distributing online "cultural tools" as well (user orientation, meta-information, availability, etc.).

Of course, this is not the end of traditional dissemination tools. While the added-value of online dissemination is to be found in being digital, in allowing multiple access points (navigation/browsing/hyper-links), in giving the user the power to produce multi-dimensional views (data warehousing/OLAP), in permitting full access to meta-information, the role of paper publications is far from being obsolete. The flop of e-books is revealing. Paper is invaluable for purposes such as documentation and reference, one-off look-ups, analysis, and promotion of statistical culture. A more serious difference is that (statistical) books are static and table-centered; online dissemination is dynamic and encourages exploration at the level of databases and data warehousing. The role for off-line digital dissemination (floppy disks, CD-ROMs, DVD-ROMs in the near future) is in-between and complementary with respect to the main two channels, internet and books.

For decades, administrative records were the main sources of data used for policy and planning for both large and small areas. These are still the richest sources of statistical data at small area levels in most countries. During the forties and fifties, however, as the reliance on sample surveys increased, survey-based estimates complemented the traditional sources because they provide more timely and cost efficient statistical data in a variety of subject matter fields. Although designed to provide reliable estimates

primarily at larger area levels such as national and provincial, increasingly such surveys are being used to meet the growing demands for more timely estimates for various types and sizes of domains.
Administrative records can provide data more frequently but suffer from coverage problems. On the other hand, sample surveys can provide information on wide-ranging topics at frequent intervals of time and at reduced cost.
Prompted by growing demand for reliable small area statistics, the research is directed towards the study of methods for small area estimation; besides, due to generalized need in formulating policies and programs, in allocation of government funds, in regional programs and so on, small area estimation has received a lot of attention in recent years.
The statistical burden on respondents is an old problem for national statistical institutes; various types of solution have been suggested to contain this phenomenon.
The methodological solution is the coordinated selection of the statistical units in sample surveys, such as rotation group sampling.
In recent years, with the appearance of new information technologies, it has been possible to by-pass the phase of interview, by getting data of interest directly from administrative files through various types of software that manage the exchange of information with limited human intervention.
However, we have to underline that a useful support for each agricultural survey is represented by the use of remotely sensed images. The use of this data source implied in the last decades the development of a methodological research areas such as: area frame sampling, statistical digital image processing and estimation based on auxiliary variables (i.e. regression or calibration estimators). Moreover this kind of data can be also used to check the quality of list frame surveys in a multiple frame perspective.
Another really important topic, strictly connected to data quality, is the minimization of the non-sampling error. This is due to the fact that the sampling variance gives an extremely optimistic view of the precision of the estimates (empirical studies verified that this parameter is approximately ten times lower than the non-sampling error).
An interesting and efficient way to reduce this error is provided by automatic data editing systems which are aimed not only at increasing the quality but also at considerably reducing the time needed for survey data processing. Thus advanced kinds of software may considerably reduce the resources needed to produce official statistics and at the same time make it possible to gain better products.
Error detection, outliers and inliers identification, imputation, global and partial non-response, are only some examples of interesting and important topics discussed during the last day of this conference.
I'm sure that these methods and techniques will significantly help us to represent agriculture through numbers.

4. Conclusions

I would like to conclude my speech by directing our attention to a relevant aspect, namely the importance of the comparability of statistics at the international level. This is particularly relevant if we focus on the developmental process in progress oriented to a globalization of the statistical information with regard to methodologies, quality and easy to use data, not only by local users, but also by users of every country of the world.

With the development of Internet it is now possible to achieve an immediate exchange of information among statistical researchers with the aim of evaluating jointly the validity and reliability of socio-economics indicators: this possibility represents a great step ahead for researchers and a warranty for the governments of each country all over the world.

In these days I'm sure that the variety and qualifications of the participants in this conference will enrich the contents of the debates and of the discussions. I hope that this work will contribute to improve the circulation and the comparison of the information between countries and the harmonization of the statistical techniques and methodologies; if so, this conference will have reached its aims.

I would like to express my best wishes to all the participants and the speakers for their work and I also wish you all a pleasant stay in Rome.

Corrado Pirzio-Biroli
Head of Cabinet of Mr. Fischler
European Commission, Rue de la Loi, 200 Bruxelles - Belgique

Mr. Chairman,
Ladies and Gentlemen,
On behalf of Mr. Fischler, who is very sorry for not being able to attend this Conference today, I would like to thank you very much for your invitation and for allowing me to make some introductory remarks to this gathering. I would, in particular, like to thank ISTAT not only for having accepted to organise this International Conference in co-operation with the Italian Ministry for Agricultural Policies, together with several international bodies, but also for hosting us in such a beautiful city.
I would also like to congratulate the Program Committee for having decided to devote some preliminary sessions of this meeting to an overview of the agricultural situation at the beginning of the new millennium. As this Conference is aimed at constituting "a departure point for the development of an information network of institutions involved in the use and production of agricultural statistics", including the "interactions with issues concerning the environment and rural development", it seems to me quite appropriate to introduce the work with a stocktaking of the current situation for Agriculture at the beginning of the new millennium.
There were many different ways to tackle this issue. I have chosen to devote my intervention mainly to some key questions in this respect and in particular to the following: is "agriculture", at the beginning of the XXI century, still sufficiently important in the economy and in the society of developed countries to justify an agricultural policy? And if so, what are the current developments in the perception of the role of agriculture in our societies, of which agricultural policies should take account?
Although we are particularly sensitive in Europe to some issues that surfaced in public debate on agricultural policy only recently, it is quite surprising to observe that practically the same questions are addressed in almost all developed countries and even in a broader context. This shows that there is not only a globalisation of economies and markets, but also a globalisation of diagnostics if not yet of therapies, as far as the agricultural situation and its associated problems are concerned.
Let me start with some factual elements providing an overview of the agricultural situation and current developments in OECD countries, on the basis of the latest reports just published by this organisation.
If we take the main indicator of the importance of agriculture in the economy, i.e. the contribution of agriculture to the total GDP, it appears that agriculture represented, on average, only 2.4% of the GDP in OECD countries in the period 1992-94, ranging from 1.5% in Switzerland to more than 15% in Turkey. Moreover, the contribution of agriculture to GDP is declining in all OECD countries as total economic output has grown at a higher rate as compared to agricultural output.
It is true that food processing represents, on average, an additional 1.9% of total GDP. This means that the agro-food sector's contribution to GDP is just over 4% in the OECD area. However, this level is still relatively low if we compare the agro-food sector with the rest of the economy as a whole.

The situation is a little bit different if we consider the share of agricultural employment in total civilian employment. Agriculture alone represents in fact 9.0% of the total civilian employment in OECD countries, to which should be added another 1.8% if we included food processing. However, these levels are significantly higher in some OECD member countries, such as Mexico and Poland (more than 25% in both cases) or Turkey (more than 45%) while they are at 8% and 4% respectively in the European Union and in the USA.

Although employment in agriculture has significantly dropped over past decades in connection with improvements in productivity and major structural changes, agriculture remains an important source of employment in most developed countries, especially if we take into account the up and downstream dimension of the food chain.

As an aside, I should mention that the agro-food sector is of major importance for the European economy as a whole. The food and drink industry is a leading industrial sector in the EU, with an annual production worth about 550 billion USD, or about 15% of total manufacturing output and it is the third-largest industrial employer of the EU with over 2.6 million employees, of which 30% are in small and medium enterprises.

Another useful indicator of the importance of agriculture and the food industry in the economy is the percentage of agricultural and food exports in total merchandise trade. For OECD countries this is about 8% on average, ranging from 0.3% in Japan to almost 41% in New Zealand. Agricultural exports remain very important for some OECD countries, notably Australia, Hungary, New Zealand and Turkey. For the United States and the European Union the share of agri-food products in total exports represents 8.3% and 9.5% respectively.

Although these percentages may be appear as relatively low compared to the total trade, it is worthwhile observing that, in absolute terms, exports of agricultural and food and drinking products are worth about 56 and 53 billion USD respectively in the US and in the EU.

In contrast to its relatively low share of GDP, employment and trade, the agricultural sector is much more important in terms of use of certain resources, in particular land use and water consumption. For the majority of OECD countries, agriculture occupies in fact more than half of the total land area and accounts for over 40% of water usage.

Finally, although the share of food in total consumer expenditure continues to decline in all developed countries, food expenditure still represents, on average, more than 12% of total final consumption in OECD countries, ranging from 8.1% in USA to almost 30% in Poland.

On the other hand, consumer patterns are changing, with consumers purchasing a greater variety of processed and prepared food products, but also being increasingly sensitive to quality and safety concerns.

Let me now come back to the first question I raised previously: do all of these factual data and observations concerning the role of agriculture in the economy show that it is still sufficiently relevant, at the beginning of the new century, to justify a specific policy for this sector?

My answer is certainly yes, although I should also mention that at least some of these arguments are becoming less pertinent while, at the same time, others, as well as some new concerns of society, are growing in importance and are becoming more relevant for justifying the intervention of public authorities in the agro-food sector.

So, if today the objective of ensuring the availability of food products may appear almost out of date in all developed economies and in an open trade framework, this is certainly not the case for other objectives which are traditionally at the heart of agricultural policy in all countries, and in particular in the European Union, namely: ensure a fair standard of living for the agricultural community, to stabilise agricultural markets and to ensure that supplies reach consumers at reasonable prices.

There are, of course, some divergences of view about the manner by which agricultural policies should tackle these preoccupations, but I think that there is a large consensus on the fact that, even today, there are strong arguments in favour of maintaining an appropriate agricultural policy in all developed countries.

In fact, what is often questioned in the international debate is not the basic need for an agricultural policy, nor the need for support to agriculture, but more often those measures that may affect a fair competition on world markets or which may isolate agriculture from the mechanism of supply and demand.

Having said that, it should also be realised that modern society demands more of agriculture than just the provision of food and feed. Above all, quality and safety of food have become increasingly important consumer concerns during the last decade.

Secondly, there are growing public concerns not only about the way in which food is produced but also about animal welfare issues, the preservation of biodiversity, agriculture's contribution to sustainable development, the protection of the environment and the preservation of the landscape.

Finally, although at the beginning of the new Millennium agriculture is not necessarily the main economic activity in rural areas, it continues nevertheless to play a crucial role in the viability of rural economies as well as in any action that could be taken to promote the socio-economic development of these areas.

As I have already mentioned, these preoccupations are particularly sensitive in the European Union, which has experienced the outbreak of different animal health crises during the last few years, where there is still a very close link between farmers and the land, and where agriculture remains a major interface between people and the environment.

However, many of these concerns are also shared by other developed and developing countries. This is because the production and consumption of food is central to any society and because it has economic, social and, in many cases, environmental consequences. It is, in fact, quite obvious that everywhere people are very concerned about the safety of the food they eat and about problems related to their quality of life.

It is mainly the role of agricultural, rural and food policies to respond to these requirements of society, should we wish to preserve their legitimacy vis-à-vis taxpayers and the ability of these policies to be consumer and civil-society driven.

Starting from food quality and food safety issues, it is obvious that consumers should be offered a wide range of safe and high quality products. This requires not only an agricultural market-oriented policy, integrating consumer's preferences and attitudes, but also a comprehensive food safety policy, which contributes to a high level of consumer health protection, while at the same time maintaining or restoring consumer confidence in the food supply and so strengthening food demand.

Ensuring food safety is not therefore in contradiction with the objective of a market-oriented and competitive agricultural production. On the contrary, not only are these different objectives quite consistent, but one could also argue that guaranteed food safety is, in most cases, a prerequisite for a product to become marketable. On the other hand, competitiveness is not only a question of low prices. The quality of a food product, the trust consumers have in it, the way in which it has been produced and marketed also play an important role in many cases.

Although the European food chain is one of the safest in the world, the European Commission is aware of the weakness of the system, which is in place in the European Union, and of the need to transform it into a more coherent, effective and dynamic food policy. This should be achieved on the basis of a comprehensive and integrated approach, involving all actors in the food chain (from farm to table), all food sectors and all partners within the EU and outside it. But it is also clear that this result cannot be achieved without an in-depth reflection on the contribution that agricultural policy can provide, in terms of re-orientation of food production, changes in methods of production and shifts in attitudes of both consumers and producers.

In this respect, it is worthwhile mentioning that, in the framework of the implementation of this approach, David Byrne, Commissioner for Health and Consumer Protection, and Franz Fischler, Commissioner for Agriculture, Rural Development and Fisheries, have called for an open debate on the Internet on the future of agriculture, food production and food safety in the EU, which will take place on the 6^{th} of June, i.e. tomorrow afternoon.

The challenge for policy makers is to find ways to match consumer demand and expectations for good quality food with an economically viable and safe food supply. This is not so easy to achieve. Many questions need to be addressed and in particular: what exactly is quality? Are consumers prepared to pay the price? Are intensive farming methods really the source of all our problems? What incentives do farmers have to shift the emphasis from quantity to quality? Can advanced technology and modern production methods deliver tasty and wholesome food? Are modern eating habits part of the problem?

Let me now turn to what we might consider as society's expectations from agriculture and agricultural policies in relation to environmental concerns and sustainable development, at the beginning of the new century.

It is not surprising that, after a long period of increase in agricultural productivity and of intensive exploitation of natural resources, there is today increasing public awareness of the environmental effects of agricultural practices and of the need to integrate environmental concerns into agricultural policies. The abandonment of land used for agricultural purposes, which is taking place in many regions mainly for economic reasons, also creates pressure on landscape and biodiversity.

But, more and more, agriculture is not only seen as one of the numerous sources of environmental pressure on natural resources, but also as a major protector of the richness of landscape and biodiversity. In this respect, farmers, while producing for the market, also play a role as the guardians of natural resources and of the diversity of the countryside and the environment.

The challenges represented by the intensification of methods of production, the abandonment of farming activities and the need to preserve the landscape raise the question of the desirable relationships between agriculture and the environment in view of promoting a sustainable model of agriculture.

Promoting a sustainable agriculture implies a management of natural resources which allows us to meet the needs of the present without compromising the ability of future generations to meet their own needs.

In the European Union, we particularly value the role agriculture plays in the conservation of the environment and in the achievement of a sustainable development,

while taking into account how negative environmental effects can be minimised. Sustainable farming has shaped the face of Europe's countryside and it is closed linked to the environment in a complex interrelationship we are seeking to promote further.
Sustainability is itself a concept having different faces. Perhaps, in the past, we have only attributed importance to the economic and social dimension of this concept. Now it is time to integrate also its environmental dimension if we want to preserve the use of natural resources for future generations. However, it would be a big mistake if at the same time we forget the need of maintaining population on rural areas and avoiding desertification.
But debate on the integration of environmental concerns into other policies and in particular into agricultural policies has not been confined to the European Union. It is worthwhile to mention, for example, that in 1992, at the Rio Summit, the signatory States adopted a series of key declarations and conventions with relevance to agriculture and forestry. In particular, the concept of sustainable development was agreed on and legally binding conventions on climate change, biological diversity and desertification have been adopted.
Agriculture is equally important to rural development. Farming is still the fabric of rural society and in many countries of the world, in particular in developing countries, it is the main economic activity in rural areas. Nevertheless, even in countries and regions where the sector is of less economic significance, agriculture continues to play an important role in socio-economic development.
"Multifunctionality" is the word that is becoming familiar in Europe to describe the fundamental link between sustainable agriculture, food safety, rural development, maintaining the landscape and the environment and what is particularly important for developing countries, food security.
As for other subjects I have already mentioned before, the multifunctional character of agriculture and land is not only increasingly recognised in Europe, including the Central and Eastern European countries, but also an issue which has been examined in-depth in several international fora, such as the FAO, WTO and OECD.
The Commission for Sustainable Development of the United Nations has also underlined the "importance and the special place of agriculture in society", due to its various functions. I am pleased to observe that, after an initial hesitancy, such a wider role and responsibility of agriculture is now increasingly understood and accepted even in the US, whatever we choose to call it.
In a world characterised by growing globalisation, resulting in rapid changes not only for agricultural production systems but also for society as a whole, there is, in fact, an increasing awareness that agriculture is able to contribute to the fulfilment of societal goals beyond simply food production.
Substantial work has already been carried out in the OECD to establish a conceptual basis for better understanding the integrated nature of the various commodity and non-commodity outputs of agricultural activity and the key supply and demand aspects of multifuctionality.
We think, however, that addressing adequately the concept of multifunctional agriculture requires further work to be done at international level, notably with regard to the collection of concrete evidence and information on the various functions of agriculture. This Conference will certainly provide useful material for reflection and for action in this area.

But, given the contribution of the multifunctional character of agriculture to sustainable agriculture and rural development, I am also convinced that it is important that a range of policy instruments should be available to governments with a view to preserving the non-commodity outputs of agriculture and taking into account each country's specific production conditions and potential, and its historical and cultural background.

We are well aware that, in the case of European agriculture, some of our partners tend to interpret multifunctionality as a synonym for protectionism and unfair competition on world markets. I do think that this would be a serious mistake. The European Union has acknowledged the benefits of world-wide trade a long time ago and it is determined to continue along this path. Multifunctionality is not a disguise for subsidies, but rather a concept that tries to take into account the various services that farmers provide and that our societies are expecting from it.

The challenge of agricultural policies in all developed countries at the beginning of the New Millennium is to succeed in meeting the expectations of civil society from agriculture, both in quantity and quality of food products, as well as in terms of development and preservation of landscape and rural areas, taking also into account the global framework for world agricultural and food trade.

The European Union has made a serious effort to adapt its Common Agricultural Policy to the new challenge. This process is under way and should be consolidated in the coming years.

In particular,

- We are progressively shifting from price support toward the far less trade distorting income support, linked to production restrictions;
- We are moving towards a competitive agricultural sector, which can gradually face up to world markets without being over-subsidised;
- We have built rural development as second pillar of our agricultural policy, recognising the multifunctional role of agriculture in the economy and in European society;
- We have started building a comprehensive and integrated approach to food safety concerns;
- We are taking measures aimed at protecting the environment and ensuring sustainable management of natural resources in agriculture.

I am not saying that all these goals have already been achieved, that the decisions already taken are sufficient to face the challenges of the future or that the instruments in place are quite adequate to the relevance of these objectives. What I would say is that this effort should be continued and even reinforced in certain areas without prejudging the need for a continuation of the reform of agricultural policies along the principles unanimously agreed in international fora.

I am sure that agricultural policies will continue to be influenced by these preoccupations during the coming years not only in the European Union but also in many other developed and developing countries.

Mr. Chairman,

Ladies and Gentlemen,

Thank you and best wishes for a successful and fruitful meeting.

Keith Collins
Chief Economist
U.S. Department of Agriculture

It is an honour to visit Italy and to have the opportunity to address this important gathering of statisticians. I thank the organizers for inviting me.

My remarks will focus on what I believe will be the most important forces that will shape world agriculture early in the 21st century. The forces I will identify are fairly obvious, but of course, my perspective on them will reflect my own experiences as a U.S. analyst. As I discuss these, I also want to comment on some of the implications for statisticians. As an economist, I am concerned about the efficiency of markets. As a public service employee, I am concerned about good public policy decisions. To perform both functions well, it is absolutely critical to have relevant, accurate, reliable, timely, and readily available data and information.

1. Performance of the global economy affects world agriculture

The first, and perhaps most fundamental, force that will shape agriculture is the global macroeconomy. As we enter the 21st century, the extraordinary events of the century we just left demonstrate the importance of a strong world economy for our future. The 50-year period following the end of World War II saw rising economic growth, expanding democracy, access to education, and advances in technology bring enormous benefits to much of the world.

The world's real gross domestic product has grown six-fold since 1950. Growth in world trade has been even more impressive, expanding almost 20-fold, as incomes rose and trade barriers fell. Rich and poor countries alike benefitted from expanding economic opportunities.

For agriculture, the story was similar. Rapid income growth boosted food demand, and adoption of new technology and rising productivity led to a tripling of world agricultural production between 1950 and 1999. The volume of world agricultural trade grew even more rapidly than production, expanding over five-fold in the last half of the 20th century.

USDA's own long-term assessment projects that world GDP growth over the next decade could average 3.5 percent per year, compared to 2.6 percent during the 1990s. Rising world per capita incomes and the addition of another 2 billion people by 2010 has the potential to profoundly affect agricultural trade over the coming decade, as the demand for food expands, particularly for higher valued foods.

According to the International Food Policy Research Institute (IFPRI), about 85 percent of the increase in world demand for cereals and meats between 1995 and 2020 will be in developing countries. IFPRI projects meat demand in developing countries will double between 1995 and 2020. To meet the rising meat demand, developing countries' net grain imports, particularly feed grains, may almost double between 1995 and 2020, and meat imports increase eightfold. Improvements in crop yields will be necessary to meet this growing demand and keep food prices steady.

But even as much of the world prospered in the last century, continuing poverty, hunger and malnutrition remain huge problems. The FAO's latest estimates indicate over 800 million people remain undernourished, mostly in developing countries. With per capita daily food availability rising, USDA projects a modest decline in the number of hungry people in the world over the next decade, but also projects that problems will become even more severe in Sub-Saharan Africa and parts of Asia. In the end, the path that global food demand takes will depend on central bankers and ministers pursuing sound fiscal and monetary policy, reducing regulatory and market uncertainty, and creating institutions that support competition.

But statisticians, too, will have a crucial role in determining how well countries and market participants fare in the expanding agricultural markets of the 21st century. All need a modern information system, which requires timely, accurate, and reliable information on agricultural production, market conditions, prices and incomes. A valid supply/demand balance sheet for traded products is essential. Credible information allows farmers to make better production and marketing decisions, supports the development of rural businesses, and is necessary for evaluating potential investments.

Yet, at the dawn of the 21st century, we still face many challenges in obtaining a modern information system. For example, we remain uncertain over basic market data in potentially major trading nations, such as China. Only last month, USDA dramatically increased its estimates of China's grain stocks, revising historical data back to 1992. Earlier, FAO had done a similar thing. These changes were made partially as a result of China's recent import levels not reflecting our estimates of their internal supplies. I cannot emphasize enough that solid data on the most basic information - what is produced, what is stored, what is traded – remains absolutely fundamental to the efficient functioning of markets and sound public policy.

2. Agriculture lags in multilateral trade liberalization

The second key force that will shape agriculture is the extent to which international trade liberalization continues. A recent World Bank study, using data from 80 countries over four decades, confirms that trade boosts economic growth and that the incomes of the poor rise one-for-one with overall growth in trade.

When the GATT began in 1947, there were 23 members and the value of world trade was about \$50 billion. By the end of the 20th century, there were 137 WTO members and the value of world trade was over \$5 trillion. Average tariffs were reduced significantly in this period; the average bound tariff on industrial products in most developed countries is now under 5 percent. The key question for the future is whether the liberalization that was achieved for manufactured goods in the 20th century can be duplicated for agricultural goods in the 21st?

Agricultural trade expanded faster that agricultural production over the past half century, but lagged behind the growth in world merchandise trade. One reason for the slow growth is that world agriculture remains highly protected. World agricultural tariffs average over 60 percent, 12 times higher than average industrial tariffs. The OECD estimates that for developed countries, government support to agricultural producers exceeded one-third of the value of OECD agricultural production in 2000. A recent USDA study concludes that eliminating all agricultural policy distortions would increase global welfare by \$56 billion and raise world agricultural prices 12 percent. All countries would benefit, including developing countries, whose food needs would decline as their domestic production expands in response to higher world prices.

The Uruguay Round brought world agricultural trade under meaningful disciplines for the first time. Since implementation started in 1995, export subsidies and tariffs have declined, and some countries have readjusted domestic support programs to be less distorting. WTO members are now negotiating to continue the process of agricultural policy reform. Negotiations on regional agreements, like the Free Trade Area of the Americas, are also underway. Yet, there is concern in many quarters that the support for free and open markets is weakening, including in my own country. The commitment of governments to achieving further progress in trade liberalization will be a key factor affecting world agricultural markets in the 21st century.

3. Will domestic policy reforms continue?

The third factor that will shape global agriculture is whether countries will make the tough choices about their own future domestic agricultural and food policies that will need to be made for the world to move forward on international trade liberalization. The good news is that many countries have made significant changes in domestic policies. This has been a response to commitments made under prior trade agreements as well as domestic economic pressures.

For example, following the implementation of the NAFTA in 1994 and the realization that import quotas and tariffs would be falling, Mexico substantially shifted its farm programs from high internal support prices to less-distorting area payments. After the Uruguay Round agreement came into effect, Canada eliminated its subsidized freight rates for grain exports, a policy that had been in effect for about 100 years. The European Union continues on its reform path to smooth the process of enlargement, meet its Uruguay Round commitments, and ease budget pressures. The combination of lower internal support prices and a lower currency value is enabling the EU is to export grains with little or no subsidy, a situation that was not anticipated even a few years ago. Argentina and Brazil have implemented major macroeconomic reforms that helped control inflation, eliminated or reduced export taxes, and lowered trade barriers under MERCOSUR. Agricultural production and exports have been booming. China has substantially lowered price supports for some key commodities to address serious distortions in its agricultural economy, partly in anticipation of joining the WTO.

The United States is now in the process of a major farm policy review as current farm programs expire next year. Some of the key issues we are grappling with are relevant for other countries, so I want to comment on our debate.

U.S. farm policy in the early part of the 20th century focused on high supported prices and production control as a way to raise low farm income and deal with persistent farm surpluses. But by the mid 1980s, several developments started pushing farm policy toward more market orientation. First, there was a national effort to reduce the Federal budget deficit, and with farm program costs running at record high levels, farm costs had to be reduced. Second, there was a recognition that the high, government-set prices and production control programs were damaging competitiveness and the environment, as well as costing consumers. And, third, we had many small farms with high incomes earned off the farm and we had large, efficient farms with high farm income, so questions were being raised about whether such farms should be supported.

Reflecting these concerns, the U.S. began to reduce government intervention and support between 1985 and 1998. Support prices were lowered, many payments were decoupled from production and prices, farmers were given more flexibility to plant what they wanted, and annual production controls for major crops were ended. Other programs, such as crop insurance and conservation, were strengthened to help farmers deal with risk and environmental concerns.

Nevertheless, our move to more market orientation has been diverted by the sharp drop in U.S. crop prices over the past several years. The U.S. Congress passed four pieces of legislation over the past three years that increased farm program spending by nearly $25 billion. And the Congress has recently decided to increase the level of funding for farm programs over the next decade, compared to what would have been available under our current programs.

The stage is now set for the debate on the next U.S. farm bill, which will go into effect starting with the crops harvested in 2003. Much of the debate is already focusing on the question of support to farmers, prompted at least in part by the WTO agreement. What form should that support take? Can support be structured to provide a safety net to offset some of the risks of farming without unduly distorting markets, penalizing consumers, or degrading the environment? These questions are germane for all countries, not just the U.S. Developing countries legitimately question whether they can have meaningful access to world agricultural markets when rich countries are heavily subsidizing their farmers. Whether rich countries can limit their subsidies and make them nondistortionary will go hand-in-hand with the success of the current WTO agricultural negotiations.

4. Can agriculture and environmental protection be compatible?

A fourth key factor now greatly affecting the debate over farm policy is the relationship between agricultural production and the environment. Public concern over global environmental and resource issues has grown over the past several decades, exemplified by issues such as deforestation, depletion of fisheries, air quality, the ozone layer and increasing greenhouse gas concentration in the atmosphere.

For agriculture, this debate ranges from water quality and availability, overuse of chemicals to food safety and animal health issues, such as BSE, *E. coli*, or foot-and-mouth outbreaks. Consumers are demanding that their food be produced in sustainable and environmentally-friendly ways, and that open space be preserved to protect wildlife and biological diversity. Reconciling these diverse demands will be a major challenge.

With respect to global climate change, agricultural activities such as conversion of forests, grasslands, and wetlands to cropland have contributed to greenhouse emissions, but may also be part of the solution. Agricultural activities can reduce atmospheric carbon dioxide through practices that reduce soil erosion and increase the organic matter content of soils.

Rising energy prices have led to a focus on agriculture as a significant user of energy, but, agriculture may also become an important supplier of energy as well. Crops and crop residues could be converted to various forms of energy, such as ethanol, biodiesel, biopower, and bioproducts. Some think that declining costs of producing fuels and electricity from renewable materials will make the 21st century the bioproduct century, just as the 20th century was the fossil fuel century.

The new emphasis on conservation and environmental issues will challenge statisticians to integrate their work with that of economists and scientists in new ways. First, we need more and better work in measuring environmental conditions and being able to link those conditions to farm activities or other causes. Second, public policy will have to address new issues that require new kinds of data and new measurement tools. For example, statisticians may be called upon to report on how identity preservation, segregation or trace back systems are working. On another front, we need cheaper, quicker tools to measure soil carbon and we need to quantify the capacity for agricultural activities to store carbon. Third, we need to improve our risk assessments and benefit-cost capabilities to develop the most effective but least-cost ways of addressing environmental and food safety issues. Such analyses require the collaboration of a range of technical specialists, and data requirements are often highly specialized and require new thinking to develop the most useful analytical tools.

IFPRI concludes that the world's farmers will have to produce 40 percent more grain in 2020 to meet global food demand. Most of that increase will come from improved yields, but pressures to increase output will continue to stress land and water resources. How do we increase agricultural output in a more environmentally sustainable way and avoid further degradation of the world's resources? That brings me to the fifth key force, which I think will be crucial in solving that dilemma-and that force is technology.

5. Technology and productivity drive agricultural change

The transforming power of technology can be seen in the U.S. transition from horses to tractors. At the turn of the 20th century, only 4 percent of U.S. farms had tractors; as we enter the 21st, the tractor is the most basic piece of U.S. farm equipment. During the same time, the area devoted to oats to feed the horses plunged from 31 million acres in 1900 to less than 2.5 million today. Land use patterns were just one of many changes resulting from machines taking the place of animals.

The last century also saw the application of science to farming, from use of hybrid seeds, improved livestock breeding, the use of chemicals, the computer and information technology. During the past 50 years, U.S. agricultural productivity grew more than 50 percent faster than the rate for manufacturing. As we start this century, the quality of technology continues to improve, whether in bioengineering, the use of satellites, communications, and the internet, or in more basic improvements, such as energy efficiency. To meet the competition of a more liberalized and competitive world marketplace, our top farmers no longer say they "can't afford" to invest in and use the new technologies; instead, they say they "can't afford not to".

The major technology challenge on the horizon is biotechnology. Many studies indicate biotech crops generally reduce pesticide use, raise yields, and improve net returns. Even though farmers appear to be benefitting from biotech crops, many consumers around the world are wary, citing potential environmental or health risks. The EU, Japan, Korea, Australia, New Zealand, and other countries are moving to mandatory labeling of biotech commodities and food.

Improved plant varieties to increase yields or increase nutritional intake are vitally important, especially to developing countries. Biotechnology holds great promise to do both, contributing to future food security and reducing the adverse environmental effects of agriculture. Realizing that promise presents a significant challenge to economists, scientists, and statisticians. To reap the benefits of biotechnology, more research is needed on the long-term environmental and food safety effects, as well as the farm-level impacts. Data collected through field trials and surveys will form the basis for this research, and statisticians will play a key role in this effort.

The combination of technological innovation and higher income growth suggests commodity markets will become more specialized and bulk commodity trade less important. Are we collecting the right data to analyze and understand these changes?

Statisticians may have many more products to keep track of as you measure the production and marketing systems of the 21st century.

6. Structural change presents new data and policy challenges

The application of technology and other factors will also challenge agriculture by changing farm and business structure, the sixth and last key force that I want to discuss.
At the end of WWII, we had almost 6 million farms in the U.S. Today, there are a little over two million. But even more remarkably, less then 10 percent of U.S. farms account for 75 percent of total agricultural sales. Over half of U.S. farms list their principal occupation as something other than farming. In other words, the U.S. farm sector consists of a small number of very large, efficient commercial operators, and a much larger number of very small operators who earn most or all of their income from off-farm sources. In between is a group of smaller, full-time farmers who earn most of their income from agriculture, yet struggle to compete with the larger, lower-cost operations. We know much about this structure because our economists and statisticians are spending more and more effort to acquire and analyze distributional data on all aspects of farm production and finances.
Along with the trend in farm consolidation, most other parts of the U.S. food system-inputs, distribution, retailing-are undergoing similar structural changes. These trends are not unique to the U.S. Globalization is perhaps an over-used concept, but certainly more and more food and agribusiness firms are operating on a global level, whether they be input suppliers, processors, or merchandisers.
Farmers and consumers are likely to have continuing concerns over concentration and related structural changes in industries that sell to and buy from farmers. The worry is that concentration may result in increased market power that lowers farm prices and raises consumer prices. Contracting between farmers and large buyers is increasing, causing some farmers to worry that local cash markets will disappear and they will become dependent on a single contractor. In an environment where the numbers of farms and commodity buyers are declining, Governments must remain vigilant and exercise strong oversight to ensure markets remain competitive.
One approach we are using to deal with our most concentrated markets is to require market participants to report market information. For cattle, where four firms buy 80 percent of the fed cattle, a new law requires these large meatpackers to report to USDA their purchase prices and all aspects of any contracts they have with farmers. The USDA then makes this information public twice per day. Farmers can use the data to better determine the strength of the market and set their asking prices. This mandatory price reporting system began April 2, and already it has offered an important lesson for statisticians. Last week, we announced that under this new program we had been under-reporting the value of a cattle carcass for six straight weeks because of a computer programming error. Farmers are now claiming this error has cost them tens of millions of dollars because USDA was telling them that market prices were lower than they really were, so they sold for lower prices than otherwise. What went wrong? While the program is now being reviewed, it appears a major problem was inadequate testing before the system was put on line. So, as statisticians, you can develop the most sophisticated mathematical analysis going, but your system may well fail if you don't provide the computer programmers-who speak a different language than statisticians-with rigid and complete specifications. And then, you must test the system extensively.

We failed to use a test database that reflected the wide range of data the meatpackers reported, and we did not conduct sufficient parallel testing-that is, running the computer system while at the same time downloading data into spreadsheets and duplicating the computer output.

7. Agricultural statistics in the 21st century

In conclusion, people living in the 21st century have the potential to be the best-fed and most prosperous ever. But to realize that potential, significant challenges in many new areas must be met-the movement toward global agribusinesses, international environmental effects, and continuing hunger in some parts of the world. Relevant and accessible information will be more essential than ever for market participants and policy officials because the challenges are more complex than ever. The 20th century gave rise to the science of statistics. At USDA, we call our statistics agency, "the factfinder for agriculture". In this new century, you, the factfinders for world agriculture, will have new ways of collecting data through incredible advances in computing power, telecommunications, satellites, and other hardware yet unimagined, as well as powerful new data base management technologies. With these advances, the ability of statisticians to reach new levels of understanding of the global agricultural forces and thereby push world leaders toward more rational and effective public policy will be greater than ever. I look forward to your future contributions.

References

Burfisher Mary, Editor (2001) *The Road Ahead: Agricultural Policy Reform in the WTO, Summary Report*, Economic Research Service, U.S. Department of Agriculture, Agricultural Economic Report Number 797.

Gibson Paul, John Wainio, Daniel Whitley, Mary Bohman (2001) *Profiles of Tariffs in Global Agricultural Markets*, Economic Research Service, U.S. Department of Agriculture, Agricultural Economic Report Number 796.

National Agricultural Statistics Service (2001) *Trends in U.S. Agriculture*, U.S. Department of Agriculture, Washington D.C..

Pinstrup-Andersen Per, Rajul Pandya-Lorch, Mark W. Rosengrant (1999) *World Food Prospects: Critical Issues for the Early Twenty-First Century*, International Food Policy Research Institute, 20/20 Vision Food Policy Report, Washington D.C..

Shapouri Shala, Editor (2000) *Food Security Assessment*, Economic Research Service, U.S. Department of Agriculture, GFA-12.

United Nations Food and Agriculture Organization (2000) *The State of Food and Agriculture 2000: Half a century of food and agriculture*, Rome, Italy.

U.S. Department of Agriculture, Office of the Chief Economist (2001) *USDA Agricultural Baseline Projections to 2010*, Staff Report WAOB-2001-1.

World Trade Organization (2000) *International Trade Statistics 2000*, Geneva, Switzerland.

World Trade Organization (2000) *Focus*, Newsletter No. 46, Geneva, Switzerland.

Issues and Challenges for Agricultural Statistics and Information Systems

Hartwig de Haen

Josef Schmidhuber[1]
Economic and Social Department, FAO

1. Introduction

Accurate, reliable, and timely information is the foundation of all empirical analysis. Statisticians play a central, but probably also a vastly underestimated, role in this context. They collect, compile, scrutinise, and disseminate information and the quality of their work critically affects the quality of the empirical analyses and thus ultimately the policy decisions that are based on these analyses.

For FAO information collection and dissemination is one of its core mandates. In fact, the Organization is perhaps the most important contributor to global agricultural statistics. FAO collects and compiles a vast array of different types of information and data series, covering agriculture, fisheries, and forestry as the main rural sectors, and, as the key domains, production, agricultural land, water and inputs, consumption, trade and markets. Over the years, FAO has assembled long time series of agricultural data and explored new ways of making them available in a more efficient manner. This process has a long tradition and in fact, could be referred to as FAO's "*traditional*" statistical data work.

In addition to the *traditional* part, there is a rapidly growing area of "*non-traditional*" data work. These non-traditional areas have become - for numerous reasons - increasingly important. The demand for better surveillance of food borne risks and diseases, for instance, has arisen from the rising concerns about food safety. Countries also continue to stress the need for better food security information and related early warning systems. Information on plant pests and animal diseases and their transmission across borders are other examples of non-traditional information work.

As both areas, the traditional and the non-traditional data work, are developing dynamically, new challenges and opportunities are arising. On the one hand more information needs to be processed, more swiftly, and often with fewer resources. On the other hand, there are new tools that are far from being fully exploited, for example remote sensing, geographical information systems (GIS) and, of course, the Internet, which enables us disseminate information faster, more effectively and efficiently.

2. The *"traditional"* areas: monitoring stocks and flows in food and agriculture

2.1 Projecting long-term trends in agriculture

Long-term perspective studies for global agriculture have a long tradition in FAO. The current long-term outlook goes to 2030 (FAO, 2000a). In a nutshell, it projects that

[1] The authors are, respectively, Assistant Director-General and Senior Economist, Economic and Social Department, Food and Agriculture Organization of the United Nations (FAO)

global food production will continue to outstrip population growth, providing more food for the world as a whole. In tandem, the number of chronically undernourished will decline, from about 826 million people in 1996/98 to around 580 million in 2015. By 2030, the number of undernourished is expected to be down to 400 million. This also means that the rate of progress will remain too low to meet the WFS targets of halving the number of undernourished by no later than 2015. In fact, the outlook suggests that it will take another 15 years to attain this goal.
The data needs for the long-term global perspective studies are very comprehensive and often pose new challenges for FAO's *traditional* statistics. These studies encompass supply/demand balances for all commodities and countries and look into the implications for the resource base, land, water, fisheries, and forestry. They use essentially all data domains in FAOSTAT and the data needs cut across various disciplines. The projections also provide a litmus test for the quality of data and, notwithstanding all efforts to ensure accuracy and consistency, they suggest that there always remains some degree of uncertainty in the underlying data source. It is therefore always important to underline that not only the projected numbers but also our assessments of the current situation are only estimates. The examples may help illustrate these concerns.

2.2 Monitoring food security and global resources

(1) How many suffer from chronic undernourishment?

FAO's estimates suggest that there are still about 824 million people chronically undernourished, the vast majority (790 million) of whom live in developing countries. These estimates are among the most important numbers that we publish and we therefore subject them to extra scrutiny, research and quality control. In a recent effort to improve the accuracy of these numbers we found evidence that the confidence intervals around the estimates of the number of undernourished could be larger than we assumed thus far[2]. Sources of uncertainty have emerged from both measuring the average calorie intake levels in a given country (DES)[3] and the possible variations around this mean. While the results of this work suggest that the estimates are robust in general, they also revealed that there is a need for more and more precise information about total availability and accessibility of food, both within countries and households.

(2) How big is the natural resource base? How much water and land is available?

Another example that underlines the need for more and better *traditional statistics* emerges from the uncertainties regarding the level and quality of our natural resources, notably land, water, and genetic resources. Over the past 30 years, world agricultural output has more than doubled and FAO's global perspective studies suggest it will increase by another 60 percent over the next 30 years. Rising crop production will remain the main pillar of output growth. About 70 percent of the increase in crop

[2] Alternative ways to assess the incidence of hunger have been scrutinised. Without delving into the details of this work, the overall result was that alternative approaches to measure and project the number of undernourished are likely to render larger margins of error.

[3] DES: Daily energy supply (average calorie intake per person and day)

production is expected to be based on higher yields, 10 percent will come from higher cropping intensity (multiple cropping and shorter fallow periods) and only some 20 percent from the expansion of arable land. While we are confident that the world has the genetic, land, and water resources to accommodate the projected growth, there is growing evidence that – in some cases - local resources are stretched to their limits. As we notice these local problems, we are also confronted with the need to analyse them more accurately which is only possible with more and more reliable information about the available resource base. A few examples may help illustrate my concerns:

China's land issue

A problem that has been haunting us since the early 1990s relates to the size of China's cropland. Over the 1990s, we accumulated evidence that revealed that China's agricultural land base is significantly larger than reported. Rather than only 95 million hectare, information derived in the context of China's first agricultural census suggested that Chinese farmers are cultivating more than 130 million hectares of cropland. The census, which was supported by many of the institutions represented in this meeting today, also provided information about other aspects of China's agriculture. It was, for instance, confirmed that the volume of crop production was in line with earlier official statistics. This means that yields must have been overestimated and should in reality be about 30 per cent lower than reported. That was an important finding suggesting that there is still a considerable gap between actual yields and theoretically possible ones. Given the size of China's agricultural economy, this is a crucial parameter for assessing China's future production potential and food security.

China's water issue

As others we are worried about a scenario where growing water scarcity could severely undermine the food production capacity of some countries or regions within countries. For instance, if China were to lose its capacity to produce enough food for its large and still growing population, some observers feel that food imports could rise to a level where they can cause significant shortages on global food markets. So the question is how severe is the water problem? We have no definite answer yet, but the available information suggests that the problem is of a local nature, within China and even more so from a global perspective.

(3) How can trade policy analysis be improved?

FAO's long-term projections also suggest that an increasing part of food will not be produced where future needs are expected to occur. Trade will be of growing importance to match surpluses and deficits at local and regional levels. For grains alone, FAO's long-term outlook for global agriculture suggests that the import requirements of developing countries will increase from about 110 million tonnes in 1995/97 to 200 million tonnes and 270 million tonnes in 2015 and 2030, respectively. At the same time, developing countries are expected to raise exports of products for which they have a comparative advantage, notably for sugar, vegetable oils, textiles or tropical beverages. Whether they will be able to do so depends – inter alia - on whether developed countries lower their level of protection and support. Developed countries' producers are

protected by higher border measures, they benefit from export subsidies and enjoy high levels of support. The OECD estimates that its members supported their agriculture with about USD 360 billion in 1999, an amount that exceeds Africa's total GDP.
These high support estimates have given rise to demands for new empirical studies which, once again, can only be undertaken with an improved statistical basis. There have been a number of initiatives to share modelling and policy data that help monitor and gauge the impacts of trade policy changes. One of the more recent and successful ones is the "Agricultural Market Access Database", better know by its acronym AMAD. It is a joint effort of Agriculture and AgriFood Canada, the Agriculture Directorate of the EU Commission, FAO, the OECD, The World Bank, UNCTAD, and the Economic Research Service of USDA. AMAD in itself may be of limited importance, but more such initiatives could promote the co-operation and co-ordination between national and international stakeholders and help us to come to grips with rising information needs. What is particularly important in the case of trade policy analyses, is that such co-operative efforts also often contribute to a confidence building process in international trade negotiations, an equally crucial and controversial area.

(4) How much is traded from country A to country B?

While we have been making progress in monitoring trade policy changes more efficiently and effectively, we seem to have still some way to go towards solid and consistent matrices of spatial trade flows. The need for such a data base becomes evident in a number of areas, one of which is measuring some of the indicators that are associated with "globalisation", which are hard to pin down without consistent trade matrixes. For example measuring the importance of preferential trade agreement is severely limited by reliable data on bilateral trade flows.
The problems associated with tracing trade flows of potentially hazardous foodstuffs are another area where more reliable information on bilateral trade flows is of crucial importance. The current BSE crisis was a case in point. Bilateral trade flows are often surprisingly difficult to trace particularly those for meat and bone meal (MBM). We found, for instance, vast discrepancies in MBM exports from the European Union compared to MBM imports of its trading partners. Yet such information is urgently needed to identify the countries at risk. This issue brings me to the next area, the non-traditional areas in which we strive to improve our information base.

3. The "*non-traditional*" information areas: the examples of food safety, food security and early warning of food shortages.

I would now like to move on to what I called the *"non-traditional"* areas of agricultural statistics, and what challenges they entail. Again, I will take an FAO perspective and focus on just three examples.

a. Monitoring food safety

The awareness of food safety issues has increased dramatically in recent years, and this not only because consumers in high income countries tend to change their preferences in favour of food quality, including food safety as their quantitative food needs are being satisfied. Additional anxieties have arisen from several other issues such as the mad cow disease, the dioxin crisis in 1999 or the outbreak of FMD and food-borne diseases due to microbiological contamination. The appearance of GMOs in food or food components has added a new controversial dimension.

What is the implication of such concerns for information systems? First and foremost, the consumers need to be fully informed; they have a right to information. In many cases, unnecessary anxieties could be avoided by informing the public about the existing science-based systems of risk assessment through which they are generally well protected against introduction of unsafe food on the market. The existing Intergovernmental Commission of the Codex Alimentarius and the affiliated expert bodies exercise a high degree of precaution before adopting food safety standards. However, information systems must also enable the consumer to be apprised of the ongoing debate among experts with regard new developments, such as foods derived from biotechnology, or of uncertainties of available methods to assess the risk of microbiological food contamination. Consumers as well as food safety regulators also need to be rapidly alerted in cases of acute outbreaks of food-borne hazards.

These needs have encouraged FAO, and other partner Agencies, in particular WHO, to elaborate, and eventually implement, a global surveillance system with respect to emerging food-borne problems. Such a surveillance system would promote the systematic collection and use of epidemiological information for the planning, implementation and assessment of disease control. At the global level it would aim to:

- inform about existing standards and risk management systems in member countries;
- determine the magnitude of existing food safety risks and monitor trends;
- identify problems at an early stage so that timely remedial action can be taken;
- determine the extent to which food acts as a route of transmission for specific pathogens, and identify high risk foods, food practices and populations;
- and assess the effectiveness of food safety programmes and provide information for formulating health policies, including preventive strategies.

A few examples may help understand the importance of such a system. With the exception of cholera, there is no obligation to report food-borne diseases internationally. Attempts to provide a global picture of food-borne diseases are usually hampered by differences in national surveillance systems, where such systems exist. In addition, the reported diseases are not presented in a uniform manner. Global surveillance of emerging food-borne problems could play a very important role in the early detection, early warning, rapid investigation and response, and help limit the extent and spread of the risk or actual contamination concerned. Information experts and statisticians are challenged to build such systems making use of the most advanced technologies.

b. Monitoring food insecurity

Another relatively new area of monitoring information flows in food and agriculture has emerged from follow-up work of the World Food Summit in 1996. The Summit concluded that there is added need to monitor food security and asked for the creation of a "Food Insecurity and Vulnerability Information and Mapping System" or short FIVIMS. FIVIMS aims to answer questions like: Who are the food insecure? Where they are located? What are their livelihood systems? What risks do they face and why

are they in this situation? With answers to these questions, development partners at all levels can combine their efforts to eradicate hunger and malnutrition.
There were four main thoughts that led to this new initiative: firstly, countries need access to up-to-date and relevant sub-national and local information with practical monitoring indicators of food insecurity and vulnerability of their people if they want to be able to target policies on these people. Secondly, modern mapping and information technologies are ideally suited for practical of policy makers. Thirdly, in order to have a realistic chance of being used, FIVIMS must build on existing food security information, which various countries have already in place. And finally, in order to ensure regular updates and responsible use, ownership of such national FIVIMS must be with the countries and managers concerned, and not with an international organization. The latter can only assist in setting-up the system, co-ordinating international harmonisation of definitions and sharing of data.
FIVIMS has indeed become a dynamic process with numerous national initiatives and an effective inter-agency working group, which is co-ordinating the various efforts. As part of the UN system reform process, it also aims to improve collaboration at the country level within the UN Development Assistance Framework. Despite the inevitable institutional challenges, FIVIMS is making significant progress based on solid technical fieldwork enhanced by new computational and communication technologies. Statisticians are urgently needed to work with FIVIMS and ensure that an interface is established with the *traditional* statistical databases. We are currently planning in international scientific conference on the methodologies of measuring hunger and food insecurity and it is our hope that statisticians will make a major contribution to this event.

c. Global Information and Early Warning System

Finally, there is another *non-traditional* information system which has been in operation for more than a decade. FAO’s Global Information and Early Warning System (GIEWS) is probably the most important world-wide system that collects and disseminates current information on possible or acute food shortages and crises. Over the years, a unique database on global, regional, national and sub-national food security has been created, refined and continuously updated. GIEWS has invested in innovative methods for collecting, analysing, presenting and disseminating information, making full use of the revolution in information technology and the advent of computer communications. The System supports national and regional level initiatives to enhance food information and early warning systems.
The main objective of maintaining a system like GIEWS is to have a tool that provides short-term forecast consistent with the information about the current situation and recent past of a country's food system. Assessment and rapid alerts for individual countries, for specific regions or the world at large are its principal outputs. The system monitors closely a wide range of factors which determine the major trends of production, consumption and markets. It also monitors and informs about factors which are the frequent cause of food shortages, including natural factors such as storms, floods, droughts, El Niño etc., as well as man-made factors such as civil strife, conflicts and major disruptions of domestic or international markets.

4. Summary and conclusions

In conclusion, I would submit for further consideration the following issues:

1. Statistics and statisticians have an unprecedented opportunity to assume a more effective role in support of policies. This opportunity is offered by the availability of modern information technologies and user-friendly information software packages.
2. There is a need to make policy makers more familiar with the opportunities of the new tools. The volume and variety of statistical information collected and compiled has rapidly increased over the past decades. And, what is more important the demands for more, more diverse, and swifter information flows are rapidly increasing. Statisticians should make every effort to show that investment in modernised agricultural information systems can have a high return.
3. There is also an urgent need to make the traditional agricultural statistical databases more usable for the newly emerging *non-traditional* information needs. Such links would help to raise the profile of the traditional systems and lower the overall cost of information. The new techniques have the potential to replace - fully or partially - some of the traditional instruments like resource-intensive questionnaires or comprehensive census approaches. Other new instruments like the Internet, help disseminate the information collected – less costly and more swiftly.
4. Managing more data, faster and more reliably calls for innovative approaches to collect, compile, and redistribute information. Such a re-instrumentation in the management of information has already been adopted in many of the new information systems on food security and bio-security. Remote sensing and GIS systems, for instance, offer important instruments supplement the traditional tools.

References

FAO (2000a), World Agriculture: Towards 2015/30, *Technical Interim Report, Rome*.

FAO (2000b), The state of food insecurity in the World, 2000, *Rome.*

FAO Committee on World Food Security: 24th Session - Report on the Development of Food Insecurity and Vulnerability Information and Mapping Systems (FIVIMS).

FAO GIEWS: The Global Information and Early Warning System on Food and Agriculture.

Emergency Prevention System (EMPRES), Livestock: Tracking trends and analysing threats from epidemics of transboundary animal diseases, background note to a presentation held on Monday, 15 November 1999.

T. Van De Venter, Emerging food-borne diseases: a global responsibility, in: *Food, Nutrition and Agriculture 2000.*

A. Pecoraro Scanio[1]
Minister of Agriculture – Italy

This Conference is extremely important because it puts together two topics of fundamental importance: agriculture and environment. During my activity in the Parliament, I have always had a particular interest in the existing interdependence between environmental and agricultural issues. In such a context, the knowledge gained from access to reliable statistical information is an indispensable and unavoidable element in order to adopt valid corrections and to minimise misrepresentations caused by uncontrolled agriculture development. I believe that the opportunity you have given us, will be very effective for analysing varied situations on a global level and, therefore, this will lead to a world-wide statistical system. I can assure you now, not only as a Minister, but also in the activity of the new Italian Parliament in connection with other European Parliaments, we will pay special attention to the meaning of a common and world-wide agricultural policy.

At FAO, I have had the chance to mention several times that Rome should be considered as the world centre of agriculture. Unfortunately, this has very often been disregarded by our own country. Italy has not given enough attention to qualifying itself as the hosting country of the food-agriculture centre of the United Nations. This can no longer be regarded as acceptable! We need a development plan for the rural areas of the world, a plan where agricultural statistics are indispensable part of our knowledge. Although famine in the world is a very important problem that needs to be faced and solved, I do not believe the objectives of agricultural planning can be limited to that alone. Instead, world rural development is a pre-condition for the environmental equilibrium of the planet. Traditionally, the FAO mission is centred on rural development rather than on food aid, which is an entirely different matter.

The objectives of this important conference, namely, the development and improvement of statistics, are essential for planning. In my opinion, the study and the monitoring carried out by the Italian government are essential and cannot be subordinated to political trends. The issue of the rural development of the world should be faced in a rational way. In this regard, we have already worked and are still working at regional level. The Mediterranean region is one of the most important. In this context, it is useful to remember that the Third Conference of the Ministers of Agriculture of the Mediterranean region was organized in Athens last week. We already decided a number of actions that, for example, are not limited to issues of free trade and tariffs; these matters, although important, risk to minimise the relevance of agriculture. The question is: how can the European Union face the problem of rural development and, therefore, the relevant problem of aids and interventions before tariffs liberalisation? In fact, this by itself is not that effective, on the contrary, we know that if we do not establish rules for globalisation, liberalisation becomes a global fraud. So, the rural development of the world is the central topic that must be faced. Before, I just mentioned the Mediterranean region, because, of course, it is obvious that the European Union is very interested in the rural development emerging in the southern side of the Mediterranean. However, if we assume for these countries the same rural development achieved in Europe in the last 50 years, it would mean first that we were thinking about spreading pesticides over

[1] This paper has been transcribed, interpreted and translated by Maria Frustaci (ISTAT – Italian National Institute of Statistics)

these areas, and then about adopting a policy of pesticide reduction, returning merely to the original sustainable production. In that way, we would risk misunderstanding the impact of such development on employment. If we wanted the entire planet to follow the European model of development, we might create billions of unemployed with no chance to do anything but to move around the world looking for a job. This is a problem that Europe recognises as urgent, because, in the next few years, the development in such areas as Magreb, where agriculture is reducing employment, will produce a further uncontrolled flow of poor people. Therefore, in order to plan better, it will be essential to have reliable and timely statistics, at national and international level, on these and other agriculture problems. Supporting this important action means putting together all our energies. In that context, the environmental impact will be fundamental, for example, in relation to the huge problems of desertification, protection of water resources and all the many problems connected to the rural evolution of the world. I believe that it is essential to use new technologies in order to gain the best possible information of the different areas and conditions in the world.

Therefore, the message that I want to communicate is that, certainly, we are supporting at national and European levels, all actions oriented to safeguarding and controlling rural development. Within the Ministries of Agriculture Council, and also in Geneva with the Director General of the WTO, I have had many opportunities to point out the need to define restrictions, as well as environmental and social rules, which are necessary and meaningful to the mechanism of free trade. For example, if we decide to eliminate immediately all the rates to pursue a fair trade with all those countries that are willing to comply with some minimum environmental and social conditions, this would have positive effects on trade and tariffs. In such context, the quality of development would be better guaranteed. At the same time, a fair competition among European farmers would be assured. In fact, importing products manufactured made with illegal environmental substances or obtained through social system such as slavery work or, as an example, the abuse of minors, could not be accepted in the European Union.

These arguments will become more and more evident and must be faced with intelligence and rationality by the international organizations to begin with FAO that, with this regard, has not always had the possibility to be supportive. Actually, FAO has not been always in a position to define seriously the objectives required by the challenge of agriculture in the world. We need maximum information and serious investment in understanding this matter because obviously, any project can be strongly affected by the inadequacy of data, particularly larger projects. Finally, I would like to remind you about the global challenge of agriculture and environment. I only need to mention the Kyoto protocol to recall those subjects that are now extremely relevant and will still be relevant at the next G8 meeting in Genoa, even as they were at the Okinawa conference. The problem of food safety, not only because of the mad cow disease, is very important because we are in an era of globalisation and every phenomenon occurring in one State has the power to produce effects all over the world. Therefore, agriculture, environment and food safety are very important issues. During the G8 meeting, environmental protocols were discussed – as they will be discussed in the future; in Okinawa transgenic products were debated. This is not only material for the news but also for political activity. Also, the quality of scientific research is a great challenge, because it is clear from all points of view that all the license system, as FAO knows, can generate wide gaps among countries and can damage food bio-diversity. The role of a world-wide statistical system is indispensable. The role of information

transparency and accessibility is another great challenge. This was debated also at the Global Forum in Naples, organised to discuss information quality and the right to have access to data, the denial of which could result in more marginalisation and discrimination, probably even stronger than merely economic discrimination. I think that this problem should be deeply analysed and debated at the international level within the context of the widest possible exchange.
Thank you very much, thanks to Istat and to all the participants. We are looking forward to the results of this conference; I am sure that they will not only be discussed by governments, parliaments and international authorities, but will also result in decisions, regulations and even the international regulations which are really the most necessary. Thank you very much indeed, I wish you a good continuation of this conference.

5 June 2001

Invited Paper Session

EVOLUTION OF THE ROLE OF AGRICULTURE IN DEVELOPED ECONOMIES

Chair: E. Giovannini

What Is Agriculture?

Susan Offutt
Economic Research Service, United States Department of Agriculture
1800 M St., N.W., Washington, D.C. 20036
e-mail:soffutt@ers.usda.gov

Abstract: In developed countries, the definition of agriculture must be flexible to accommodate the range of questions that society has about farms, rural people, and the environment. Adoption of the farm household, rather than the farm business, as the basic unit of observation in agricultural surveys and censuses supports this flexible definition of agriculture by capturing data on social and economic characteristics.

Keywords: farm census, farm household

1. Agriculture in Developed Countries

How do developed countries define their agricultural sectors? The answer is that it depends on what is at issue. Is it food security? If so, then the definition has mostly to do with the relatively small number of commercial farms on which the bulk of food is produced. Is it land use? If so, the definition would have to include many landholdings on which little is produced for commercial sale but which account for a good portion of the rural land base. Is it rural poverty? Then an accounting of land holding or land use may be of less concern than an assessment of the wellbeing of people living on the land. In such circumstances, the definition of agriculture takes on a plasticity that can be troubling to those in the analytical community.

Twenty years ago in the United States, agricultural economists and statisticians expended much effort trying to define agriculture in a precise way. A now-familiar argument arose about deleting from the ranks of officially counted farms those that essentially produced nothing. A United States Census Bureau official wrote, “The present concept includes a large number of units that make little contribution to total agricultural production and have little relevance to agricultural programs. To the extent that the individuals and families involved with these small units present problems affecting the public welfare, these are not problems within the framework of agriculture….” (Taeuber, 1972). The rationale was that these farms made no material contribution to commercial food production, and so, from the standpoint of national income accounting, nothing was lost by their exclusion. Analysts decried the use of sector averages, such as per farm output, that masked significant differences across farms. Some misuses of farm definition were identified as “odious,” to wit, the eternally popular “statistic” that purports to show how many people each single farmer “feeds” (American Agricultural Economics Association Committee on Economic Statistics, 1972). All this imprecision was just too much for the analytical community to bear.

It should be sobering to realize that none of this hand wringing had any effect on the willingness of American politicians, farmers, or citizens in general to adopt a more precise view of agriculture. Against all professional advice, the press and candidates for political office continued to appeal to the stereotypical view of farming, of a diversified operation of a certain size, worked by a farm family, feeding themselves and their fellow citizens with the fruits of their hard labor. Beyond fulfilling a romantic notion of farming in American society, sticking to an outdated view of farming had some economic advantages. A number of Federal programs used the farm counts to dole out funding, so any change in the definition of a farm would in turn require change in the distribution of funds. So the definition of a farm had a practical application that provided a powerful incentive to remain with the status quo. Sentiment and practicality, then, together ensured that the definition of a farm would remain essentially unchanged. For the United States, a farm is any place that has or normally would have $1,000 of agricultural product sales in a year, and, by latest agricultural census count, there are about 2.1 million of them.

2. The Social Reality of Agriculture

How do analysts who are dedicated to accuracy cope with such a muddle? Clearly, arguing that agriculture is not what most lay people believe it to be is an uphill battle. Still, there are those who fight it every day. But there is an alternative that satisfies both the desire (and obligation) of analysts and statisticians for precision and also the requirements of societal discourse. That solution is, of course, to allow for multiple answers to the question, "What is agriculture?" without insisting any one reign supreme. Each individual answer, then, can be tailored for accuracy suited to its purpose.

Now this is good news insofar as it not necessary to reach consensus on a unique definition of agriculture. The experience of the past twenty years teaches that it is probably an impossible task anyway. But it could be bad news for public agencies charged with the measurement, characterization, and analysis of agriculture in all its dimensions. It could be bad news if the agencies insisted on the purity of a single definition in defiance of society's clear preference for multiple identification. But it could also be bad news if these agencies fail to listen to the full range of questions to which society demands answers and that require the information contained in statistical measurement and analysis of agriculture. After all, the point of a statistical information system must be to provide government and citizens with information that is relevant to the decisions they must make. Of necessity, this flexible approach requires modification to data collection schemes as interests and circumstances of agriculture change. There can be no static framework for measuring and assessing the condition and contributions of agriculture and the people involved in it.

To provide a socially relevant and responsive statistical system for agriculture, it is well to recognize there are two main branches of public interest in agriculture. One is primarily social, it is concern with the wellbeing of people who live on the land, whether they are commercial scale producers or not. This social concern extends to important non-market outcomes, most notably environmental quality, which is determined by the ways that people manage the land. The maintenance of environmental quality on farms and more importantly off-farms dictates this interest. Demographic data and data on environmental quality, then, become part of the profile of agriculture. The other branch is a more traditional one that is concerned with the economic organization of the farm and food sector and its contribution to and inter-relationship with the rest of the national economy. Here, there are familiar questions to do with vertical integration and other forms of business organization that seem to be alternatives to the traditional setup of farms seen as atomistic business units run by independent sole proprietors.

The public role in gathering market intelligence is often justified as necessary to level the playing field on which numerous small farmers face larger firms as both buyers and sellers. Collection of market data on prices and quantities in exchange is a key component of servicing this facet of information needs for agriculture. Here again, there was recognition twenty years ago that price data might not truly reflect the value in exchange because of the use of contracting and other inter-firm agreements. At that point, there was a strong feeling that firms ought to be made to turn over the relevant proprietary information to allow agencies to get an accurate depiction of market activity. That issue has never been fully resolved. Mandatory price reporting for livestock in the United States is at present stymied by the confidentiality restrictions that compel aggregation of reporting by a few large meatpackers. It is hard to think of an example of such a direct confrontation over public commodity market information occurring at any time over the past two decades. These circumstances may call the political question: will legislators compel full disclosure of price information even at the expense of revealing the identity of individual packing firms? It remains to be seen, but whichever way the question is decided, it will have significance as a precedent in other agricultural market settings, where consolidation and vertical integration continue.

3. A Flexible View of Agriculture

Happily, it turns out that servicing a complex view of agriculture is entirely feasible. But doing so does require revision of traditional views of the uses and scope of agricultural and farm statistics. The focus here is on the demographic and economic characteristics of what in the United States are defined as farms, rather than on the description of market structure and transactions. How can statistical and analytical response adapt to multiple definitions of agriculture?

The most significant change in outlook required by flexibility is the adoption of the farm household as the relevant unit of observation. Traditionally, the farm business has been the focus of attention by public statistical and economic agencies. However, many of the questions society has about agriculture can only be answered by understanding the behavior of the farm household, of which the farm business may be an important part or it may not. Twenty years ago, a special committee of the American Agricultural Economics Association wrote that, “in any data system the common building block for all other data the system is capable of producing is the basic unit of observation. Farming has become a heterogeneous and functionally dissimilar set of activities and processes. It is no longer possible to use the farm as the basic unit of observation” (American Agricultural

Economics Association Committee on Economic Statistics). As suggested earlier, the views of these learned individuals were at odds with the social construct of agriculture. The farm continued as the basic unit of observation in national farm surveys and censuses. Now, what is compelling about adopting the farm household as the basic unit is that it does not require that the underlying farm business concept be abandoned or that some farm units be excluded because of size. Instead, the information on the farm business, however modest its size, is simply incorporated as one activity within the unit of observation, the farm household.

The insight about the value of using the household as the basic unit of observation is quite obvious in the literature on economic development, which has always used the household framework to understand the behavior of those engaged in agriculture. For example, decisions about natural resource management may well be related to household goals that are not directly reflected in farm business accounts. It may be that the expectation that children will take over the farm affects decisions about resource management. Moreover, because many farm families have significant off-farm income (and negative farm income), it is not possible to understand their management of the farm business without knowledge of other sources of income and assets. Viewing the household as the relevant unit dictates expansion of collection of data beyond the farm business. Public data collection agencies may be reticent about this expansion, as it can be viewed as intrusive by those interviewed (and also by the interviewers, who may be reluctant to press for this information). But, if these data are not collected, then public agencies will fail in their role in answering society's questions.

With the farm household as the unit of observation, the next adjustment to traditional treatment of survey data is to use it in cross section, not simply in aggregates or sector averages. Here, the challenge is to characterize the heterogeneity that exists across farm households in a way that is useful for policy discussion. Viewing survey or census data in cross section allows identification of key sub-groups of the farm population. If the issue is food security, particular attention can be directed to those households with commercial scale farming operations. If the concern is rural poverty, those households with farm and off-farm income below a threshold level can be identified. Stratifying the households geographically can be useful, especially when sensitive environmental areas need to be singled out. What is key about viewing the data in cross section is the recognition that the political process is less concerned with efficiency than it is with equity. Identifying those who gain and those who lose from different policies or programs is a key use of the public database on agriculture because of the influence on political decision making. The traditional lens for looking at the impacts across farms of agricultural policies is commodity specialization. When the effects of traditional commodity programs are at issue, this breakdown of farming operations may suffice, but for other questions it will not. For example, concern about the wellbeing of farmers nearing or in retirement requires assessment of the age distribution of farmers along with information on financial status, including, importantly, farm and non-farm resources. The perennial question about policy effects on "large" versus "small" farms in the United States reflects this political interest in the distribution of benefits. Survey or census data viewed in cross section may show the extent to which farm size is the issue versus some other characteristic, such as financial efficiency.

4. An Illustration from the United States

The assertion that it is indeed feasible to cope with a flexible view of agriculture is based in the experience of the Economic Research Service (ERS) at the United States Department of Agriculture. In partnership with the National Agricultural Statistics Service (NASS), the agency that conducts surveys as well as the census of agriculture, ERS analysts have developed a typology categorizing farms into more homogeneous groupings than classifications based on the traditional division by volume commodity sales alone. The typology is based on annual farm survey data collected in the Agricultural Resource Management Study by NASS and ERS. As such, the data provide a statistically reliable picture of American farming. The ERS typology is thus an effective tool for understanding farm sector behavior and well being and developing appropriate policy and an improvement over most policy models that rely on non-survey constructions of "representative farms."

Since the 1930's, American farming has been transformed by technological and economic opportunity. Advance in mechanical equipment not only allowed crop specialization to take advantage of scale economies, it also saved labor, releasing farmers and farm workers to higher-paying jobs in industry. Urbanization, the growth of suburbs, and the development of rural economies allowed farm families the opportunity to live on the farm but also work in non-farm jobs. As average farm size increased, farm numbers fell steadily over the decades following World War II. Today, farm numbers appear to have stabilized at just over 2 million. But what the country does not have is simply a shrunken 1930's-style farm sector. Most farms today are small and account for only a modest share of agricultural production, even if they control three-fourths of the country's farmland. The largest farms operating on the other quarter of farmland grow more than 60 percent of food that enters commercial channels. Almost two thirds of all farm operators do not regard farming as their main occupation, but rather live on farms as a retirement or residential lifestyle choice.

The ERS farm typology that describes the transformed sector is based on the occupation of operators and the sales class of farms combined. It identifies five groups of small family farms (sales less than $250,000).

- **Limited resource** Any small farm with gross sales less than $100,000, total farm assets less than $150,000, and total operator household income less than $20,000. Limited-resource farmers may report farming, a non-farm occupation, or retirement as their major occupation.
- **Retirement** Small farms whose operators report they are retired (excludes limited-resource farms operating by retired farmers).
- **Residential/lifestyle** Small farms whose operators report a major occupation other than farming (excludes limited resource farms with operators reporting a non-farm major occupation).
- **Farming occupation/lower-sales** Small farms with sales less than $100,000, whose operators report farming as their major occupation (excludes limited-resource farms whose operators report farming as their major occupation).
- **Farming occupation/higher-sales** Small farms with sales between $100,000 and $249,000 whose operators report farming as their major occupation.

In addition, there are three categories of other farms, considered large in that their sales exceed $250,000. This threshold is admittedly arbitrary, ERS choose $250,000 at the suggestion of the National Commission on Small Farms.

- **Large family farms** Farms with sales between $250,000 and $499,999.
- **Very large family farms** Farms with sales of $500,000 or more.
- **Nonfamily farms** Farms organized as nonfamily corporations or cooperatives, as well as farms operated by hired managers.

This typology now forms the basis for disaggregating ERS reporting on farm household and business performance and will be used to evaluate the impacts of proposals for change to agricultural legislation.

The typology permits any number of interesting comparisons of key farm characteristics across groups. But the perhaps most relevant for farm policy discussions is farm household income. Looking only at the 1999 national average farm household income of $64,347, the survey shows that 90 percent of it comes from off-farm sources. However, disaggregating using the typology shows very clearly how dependence on farm income varies by farm type. Only for households operating very large farms does income from the farm business contribute more than 80 percent of total income. For large farms, farm income accounts for 60 percent and for higher-sales small farms half comes from farming. The remaining small farm households derive virtually all income from off-farm sources. Off-farm income, therefore, is as important or more important than farm income to the wellbeing of most of America's farm families. There is a political argument that farm families ought to be able to earn their living entirely from farming, not a very realistic proposition. So, it is important to recognize that the ability of any farm policy instrument to affect farm household wellbeing is limited for a large portion of the small farm population. The importance of a vibrant off-farm economy cannot therefore be overstated.

The data on household income also show distinct differences in levels compared to U.S. average household income. As noted, the average farm household income in 1999 was $64,347, about a third higher than the average for all U.S. households. But, again, this average masks significant variation. On the one hand, the average household income for limited-resource farms of $9,534 lies below the poverty level but that for the very large family farms ($201,206) is more than three times the national average. On smaller farms whose operators main occupation is farming, the higher-sales group's total income is just above the national average but the lower-sales group lies just below, as it does for retirement farms. Residential/lifestyle farms have negligible or negative income from farm but household incomes above the national average.

These comparisons of farm household income across typology groups demonstrate one use of survey data that emphasizes a cross-sectional view and the value of using the household as the basic unit of observation. For more information on the typology and the ways it can be used in analysis, see the most recent ERS report on the family farm (Hoppe, 2001).

5. Challenges ahead

The introduction and use of the ERS farm typology in reporting and analysis of agricultural issues has helped orient policymakers and the public to the implications of the diversity of farming in the United States. The typology categories, although arbitrary, appear to have intuitive appeal to users of ERS analysis and do not compel them to abandon the many small units they believe are truly farms. By not forcing farms into a one-size-fits-all definition, the typology facilitates discussion about social and economic dimensions of agriculture. Ideally, all developed country agricultural survey and census systems would adopt the farm household as the unit of observation, and accounting record keeping systems would likewise be expanded. If such data were available, it would provide a much-needed empirical basis for international dialogue about the maintenance of the wellbeing of the farm sectors across developed countries.

References

American Agricultural Economics Committee on Economic Statistics (1972) Our obsolete data systems, *American Journal Agricultural Economics* , 54, 867-875.

Hoppe R.A. (2001) *Structural and Financial Characteristics of U.S. Farms: 2001 Family Farm Report*, U.S Department of Agriculture, Economic Research Service, www.ers.usda.gov.

Taeuber C. (1972) Future structure of census data relating to agricultural and rural people, *American Journal of Agricultural Economics,* 56, 909.

New Policy Challenges: Requirements for Data Collection and Analysis

Gérard Viatte
Director for Food, Agriculture & Fisheries, OECD
2 rue André Pascal, Paris, France, e-mail: gerard.viatte@oecd.org

Abstract: Agricultural policies are becoming more complex, and increasingly interlinked with broader policies. This creates new demands on the agro-food statistical system, at the domestic and international levels, in particular the need to quantify the impact of policies. New institutional frameworks need to be developed involving all stakeholders.

Keywords: agricultural policy; agricultural trade

1. Introduction and Summary

The evolving nature of the socio-economic parameters of the agro-food sector, and of its relation with the broad economic and societal context, have created new policy challenges which are increasingly complex and sometimes difficult to quantify. At the same time, as the policy debate has involved an increasing number of stakeholders, the need for an objective basis for the elaboration and implementation of policies has become more acute, hence the search for relevant data bases. In addition, the process of globalisation has given an international dimension to practically all aspects of the agro-food policy debate, whereas some issues are definitely of a regional or local nature. This has created tensions between the need to adopt an international approach on the one hand, and the recognition of the site-specific nature of many agricultural issues on the other hand.

OECD has been over the last few decades, and is still today, at the centre of these complex, and sometimes conflicting, developments. First, as an organisation aiming at providing sound and objective analysis for policy elaboration, it is involved in developing new economic and statistical concepts which can clarify and substantiate the policy process. Second, for an intergovernmental organisation, this "policy-relevant analytical work" needs to have an international dimension, which presents specific requirements for data collection and integration. Third, as OECD is dealing with practically all economic and social issues, it applies an "integrated" approach to all its work, which means that its statistical activities are also implemented on a broad base, emphasising the relations between the various sectors. Fourth, although OECD is an intergovernmental organisation, it has increasingly recognised the need to set up a participatory process with all stakeholders, including in the agro-food sector: even in the field of statistics, the involvement of private stakeholders is today necessary, at least for the selection and the interpretation of the data.

In this general context, the changing requirements for data collection and analysis can be illustrated by some major examples of OECD activities, it being understood that these examples are not exhaustive. As the OECD work programme is determined by its Member countries by consensus, it can be considered as a good indicator of the evolution of the policy agenda in developed countries. For each example, a few conclusions are drawn which have a general significance for data collection and analysis.

2. Calculating the Level of Support to Agriculture

Since the early eighties, the economic and political difficulties arising domestically and internationally from the "traditional" agricultural policies implemented in OECD countries since the second World War, have led to the need to analyse their impact in a systematic and objective manner. The first requirement was to calculate the level of support given to agriculture in order to increase transparency and to create a basis which would help to assess the impact of agricultural policies on the domestic economy, and on international trade. OECD was requested by the Ministerial Council to develop the necessary instruments and to that effect selected the concept of "Producer Subsidy Equivalent", originally developed by Tim Josling.

The advantage of this concept was that it was comprehensive, covering all types of support, in particular that resulting from market price support and border protection on the one hand, and that resulting from various types of budgetary payments on the other hand. This advantage of comprehensiveness was complemented by the fact that it was possible to disaggregate it into its various components. OECD applied progressively the same methodology to all its Member countries, and to a number of non-Member countries. As a result, it can be said that the agricultural sector is today the economic sector for which the level of support is best documented, in spite of the inherent difficulties and complexities of agro-food policies (or perhaps because of these difficulties, which justified a major effort by the international community).

Of course, the methodology evolved over time, and a number of specific problems had to be overcome when the methodology was applied to a wider variety of countries and to a wider range of commodities. Let us just refer here to the difficulties in applying the methodology to Central and Eastern European Countries before their economic reforms, or to the fact that some important sectors, such as fruit and vegetables, are only partially covered at the moment.

The definition of PSE evolved slightly over time, and the concept was renamed a few years ago to "Producer Support Estimate", in order to better reflect its economic meaning. It also became necessary to complement this concept by other instruments. The present definition of PSE is the following:

Producer Support Estimate (PSE): an indicator of the annual monetary value of gross transfers from consumers and taxpayers to agricultural producers, measured at the farm-gate level, arising from policy measures that support agriculture, regardless of their nature, objectives or impacts on farm production or income. The overall PSE monetary value depends on the size and structure of a country's agricultural sector, as well as on the monetary unit used. Support (PSE) expressed in relation to the number of farmers or area of farmland is influenced by differences among countries in factor endowment and the number, type, and size of farm holdings. By contrast, support expressed as a percentage of gross farm receipts (%PSE) shows the amount of support to farmers, irrespective of the sectoral structure of a given country. For this reason, the %PSE is the most widely used indicator for comparisons of support across countries, commodities and time.

For an international organisation such as OECD, the major advantages of the PSE approach for the analysis and evaluation of agricultural policies are the possibilities to:

- Make comparisons between countries, which reflect a wide discrepancy among them

- Make comparisons among commodities, as the level of support is extremely difficult to assess in this respect.

- Make comparisons over time, as the increases which occurred since 1998 after a period of about 10 years of progressive, although limited, reduction, were a very important and worrying political signal.

- Analyse the components of support. This objective is now a priority interest, as agricultural policies became diversified, reflecting, in particular, a limited shift from the traditional form of market price support to more direct forms of support which are deemed to be less distorting. The share of market price support has indeed declined from 82% of total support in 1986-88 to 72% in 2000, but it clearly remains the dominant form of support.

The creation by OECD of this consistent set of data has not only been extremely useful for the general analysis of agricultural support, but it has allowed the development of concepts which are "policy-operational". The most striking example is the "Aggregate Measurement of Support" (AMS), which is the negotiating instrument used in the URAA to measure agricultural support and to set up disciplines for its reduction. The data base developed for PSE has also helped the process of tariffication used in the URAA to determine, through a single instrument, the level of protection and the commitment to reduction. The comparisons between the developments in PSE and in AMS provide by themselves very interesting analytical insights. For example, OECD has recently calculated that 60% of total support, as measured by PSE, is not covered in the AMS as defined by the URAA. This means that the majority of support is not subject to the reduction commitments of the URAA, because it belongs to the "blue box" or the "green box" which include "authorised" forms of support considered as not distorting or minimally distorting. However, when these forms of support become predominant, it is difficult to pretend that they have no, or only minimal impact, on production decisions and therefore on trade.

Many lessons can be drawn from this OECD experience on PSE calculations and monitoring. In relation to data collection and analysis, I will limit myself to the following four :

i) The development of a policy-relevant indicator requires assembling, in an analytically consistent framework, a large amount of data from different origins and nature, ranging from "classical" data on production to more difficult data on domestic and world prices, and to data originating from non-agricultural statistics, such as budgetary data.

ii) Even if the political and economic characteristics of various countries are naturally different, it is necessary to achieve the highest possible degree of consistency, in order to allow for international comparisons, and for international aggregation (for example, the calculation of the aggregate level support to agriculture in OECD as a whole of $327 bn in 2000, i.e. 1.3% of OECD GNP, is a "powerful" instrument from a political viewpoint).

iii) A strong network of co-operation between national authorities and an international body is necessary in order to apply, in a consistent manner, a complex methodology requiring very different data sets. "Quality control" has to be implemented very strictly by the international body, which should also provide a framework for a continuing exchange of views and experience, and for a dynamic development of the concept in order to reflect political priorities, while maintaining continuity. The political strength of the PSE lies in the fact that it has been formally approved by all countries in the context of the OECD Committee for Agriculture.

iv) A complex quantitative indicator such as PSE needs to be integrated into a broad analytical framework, using other quantitative data and, above all, a "qualitative" analysis reflecting both political and economic developments. For this reason, OECD never publishes PSE figures in isolation, but only in the context of indepth reviews of agricultural policies of specific countries, or in the context of its annual report on "Monitoring and Evaluation of Agricultural Policies".

3. Measuring the Impact of Policies

The above description and evaluation of PSE make it clear that this concept does not aim at quantifying the impact of policies, although it permits one to distinguish between the various categories of policies - a first step in the direction of a more refined analysis. In fact, OECD has recently revisited the disaggregation of PSE into its various components, in order to better reflect recent policy developments.
In order to get a more systematic analysis of the impact of various types of policy, which could also be formally accepted by all countries, OECD embarked upon an ambitious process to develop a "Policy Evaluation Matrix" (PEM). This process, which involves co-operation with national experts, and continuing surveillance by the Committee for Agriculture, started in 1998 and led to the publication in May 2001 of a first report entitled "Differences in Production, Trade and Income Effects Among Farm Support Measures", which applies the PEM approach to the crops sector.

The PEM analysis aims at quantifying the differences in the impact on production, trade and farm income, of the various types of support policy. The analysis has focussed on five "stylised" crop support measures: market price support and four categories of budgetary payments based on, respectively, output, variable input use, planted in the current period and areas planted in a historical period. Although "stylised", this classification is based on existing policies and the crucial question for the analysis was: “How do the various forms of budgetary payments differ from market price support and from each other in their market and economic effects?”

In developing an approach such as PEM, the question is not only to define the various types of support of which one wants to measure the impact, but also to define the various groups of economic agents on which one wants to measure the impact of support. PEM has distinguished between five groups: 1) consumers, 2) taxpayers, 3) landowners, 4) suppliers of non-land farm-owned factors (mainly farm household labour) and 5) suppliers of purchased inputs. Consumers and taxpayers pay the monetary costs of support while landowners, farm households and input suppliers are the beneficiaries.

The general conclusion of the PEM analysis is that all ways of supporting crop producers may distort production and trade. Market price support is indeed a relatively inefficient and trade distorting way of supporting farm incomes. Budgetary payments based on output or on variable input use were also found to be potentially as inefficient and trade distorting.

The PEM activity leads to the following remarks in terms of statistical requirements and analysis:

i) The measurement of the various types of support originates from the PSE data base. Hence, a given database, which originally aims at giving a "photograph" of the level of support, can serve as a basis for an impact analysis. For OECD, this connection is important, as it allows for analytical consistency and efficiency in the use of statistical resources, while permitting a major breakthrough in addressing the question which is now at the centre of the policy debate: Are some forms of support really less distorting from the economic and trade viewpoints?

ii) The fact that a model approach, based on recognised data bases and validated by all Member countries of an international organisation such as OECD, is very important for domestic and international discussions. However, the PEM approach has been resource-intensive, and its application to other commodity sectors, such as the dairy sector, will require further efforts. Even though the methodological framework is now well established and generally agreed, the collection of data for such a complex sector as dairy will be resource consuming.

iii) So far, the PEM approach has been mainly used for measuring the impact of various types of policy measures on production, trade and income. But it is tempting to widen the matrix by aiming at measuring impact on employment and on the environment, which are two highly policy-relevant parameters. Some attempts have been made in this direction, but they have been limited to specific cases or countries. The limitations in the availability of relevant statistical data, and the difficulties in interpreting them in a meaningful economic manner, constitute two obstacles. They could theoretically be overcome in the future, but the practical implementation on a large-scale and comparative basis will require time and effort.

iv) A solid analytical system, developed with the help of highly-qualified professionals, based on recognised data bases and economic parameters, and validated through a critical clearance process at the international level, can provide quantitative results which are highly relevant for the policy debate, even on very difficult issues such as impact analysis. The permanent co-operation between the policy analysts, the econometricians and the statisticians, is a prerequisite for any such development. The co-operation mechanisms used for PEM were twofold: a network of professional experts, geared by a "scientific approach", and the official discussions in the OECD Committee for Agriculture framework. These two processes worked constantly in an interactive mode. This may well serve as a model for the development of such complex policy-relevant quantitative analysis.

4. Assessing the Impact of Trade Agreements

The analysis of national trade regulations and of international trade agreements has always been a major challenge for policy analysts and for economists, in particular in the agro-food sector which was characterised by an extremely complex set of regulations. The tariff structure was generally much more complex for agro-food products than for other commodity groups, and the array of non-tariff measures was certainly much wider.

The URAA was a major breakthrough, as it included for the first time the agricultural sector in a worldwide trade agreement. The agreement, which included multilateral commitments for the reduction of import tariffs and export subsidies, and of domestic support, represents a milestone in international agricultural negotiations, even if its practical impact on trade has so far been rather limited.

It could also be hoped that, by its very existence, the URAA would lead to the elaboration of a much more comprehensive and consistent database for the analysis of agricultural trade developments and measures. This had indeed been achieved, but to a limited extent only, and new problems for data collection and analysis have arisen. Mention has already been made (section 2 above) of the need to complement the data on AMS by monitoring the wider analysis of PSE in order to get a full picture of domestic agricultural support.

As far as market access is concerned, the tariffication process has certainly clarified the situation. However, the tariff structure remains extremely complex and difficult to analyse. The fact that some tariffs have been set at extremely high levels, which are clearly prohibitive, also makes analysis of the impact of the agreed tariff reductions irrelevant in such cases.

The major difficulty in the analysis of the URAA's impact in terms of market access derives from the creation of tariff rate quotas (TRQs). Countries were obliged to provide a minimum level of import opportunities for products that were previously protected by NTBs, while also ensuring that import volumes that took place prior to the Agreement could continue to be imported, by establishing tariff rate quotas (TRQs). This import system established a quota and a two-tier tariff regime for affected commodities. A lower tariff applies to imports within the quota while a higher tariff applies to imports exceeding the quota. Countries are not obliged to actually import the stated volumes. As of May 2000, a total of 37 countries, including all OECD Member countries except Turkey, with approximately 1370 individual TRQs committed to this system. These countries also agreed to notify the WTO each year about their progress in meeting market access commitments.

Since the implementation period for developed countries is now over and negotiations on further market access reforms are underway, many are interested in assessing the effects of the Agreement on tariff levels and market access, along with assessing potential effects of further reforms. Data to assess the success of the Agreement are not readily available, especially in a form that is easily downloaded, searched, or manipulated. Furthermore, the task of collecting the data and placing them in a format that is amenable to further analysis is a costly and time-consuming undertaking.

Recognising this, several national and international organisations, including the FAO and OECD, agreed to pool their resources to develop a database that would be freely available to the public and easily accessible (AMAD, Agricultural Market Access Data Base). The co-operators agreed to participate in this activity because they all require accurate data on tariffs, tariff rate quotas, and trade data to enable research on market access conditions in the international agriculture and food sectors.

The purpose of this database is to provide a common data set on tariffs, TRQs and imports, as well as tools for researchers, policymakers, and others to use in analysing levels of tariff protection in agriculture among WTO Members. The development and use of a common data set can assist in improving international transparency of agricultural trade, as covered by multilateral rules and disciplines.

AMAD is a very good example of the creation of a new database to respond to a new policy environment. Some lessons can already been drawn from it:

i) An international trade agreement has established a new policy framework, and in particular a new instrument (the TRQs), thereby creating the need for a relevant data set.

ii) When the raw data are extremely complex and originate from different sources, it is necessary to create a systematic framework in order to elaborate a meaningful data set which can in turn be used for analysis.

iii) The development of this database is institutionally independent from the negotiating process (WTO itself is not a member of AMAD, although it is fully informed), but it aims at helping the analysis of the impact of the existing agreement (the URAA) and of various scenarios on market access for the ongoing WTO negotiations which started in 2000.

iv) The creation of AMAD is the result of an original co-operation mechanism between the competent agencies of a few countries (Agriculture and Agri-Food Canada, the EU Commission, USDA/ERS) and of a few international organisations (OECD, World

Bank, UNCTAD), with OECD acting as the secretariat of the Steering Committee and as manager of the website. This shows that innovative co-operation mechanisms can be developed, even among institutions which have not exactly the same legal nature.

5. Responding to Societal Demands

5.1. Environment

Beyond producing food and fibre, agriculture's relations with environment, rural development and other broader socio-economic dimensions are today playing a major role, and some of these relations are becoming major factors for the design of agricultural policies.

As a consequence, data requirements for the analysis of these complex relationships are becoming politically and socially more important, require a broader perspective than the traditional agro-food approach, and therefore more intensive co-operation between the "agricultural statisticians" and other statisticians.

The most striking example is the mutual relation between agriculture and the environment, including both the positive and negative impacts of agriculture on the environment, as well as the incidence that environmental trends may have on agriculture. The need to develop a database which could throw light on these complex interrelationships was recognised by OECD as a priority many years ago, and gave rise to an ambitious programme to develop agri-environmental indicators (AEI). As this programme, and in particular its recently published results, will be presented in this volume by my colleague, Kevin Parris, I will limit myself to three personal comments which I drew from this ambitious and successful programme:

i) The development of an agreed database is absolutely necessary to provide an objective basis for the understanding of the complex linkages between agriculture and the environment, for the assessment of existing policies, and for the further elaboration of well-targeted policies.

ii) Even if many of the issues are "site-specific", international co-operation is essential to elaborate consistent methodologies, and to arrive at results which can be compared internationally to a satisfactory degree.

iii) Agri-environmental indicators have to be developed in order to respond to the specific needs of the sector, in particular to the "problematique" of agri-environmental policies. At the same time, they should be compatible with a broader set of environmental indicators, so that they can constitute one of the major segments.

iv) AEI may also be input into the wider framework of "sustainable development indicators". To this end, they should incorporate data which cover the three intrinsic aspects of sustainable development: economic, social and environmental. This would be particularly useful, as agriculture, and natural resources in general, are key parameters for sustainable development.

5.2. The "Multifunctionality" of Agriculture

The fact that agriculture fulfils other functions than producing food and fibre has always been recognised, but the concept of multifunctionality has been used more recently in the context of domestic and international policy discussions. In order to avoid premature and unfounded policy conclusions, OECD countries agreed on the need to develop a conceptual framework which would allow for an objective and systematic analysis. They recently agreed on the report prepared by the Secretariat and published in May 2001. This report, which constitutes the first phase of the OECD work programme on multifunctionality, does not aim at elaborating relevant data bases, nor at quantifying the issues. However, one can already identify the statistical data and analysis which will be required, to respond to the following three fundamental questions which are highlighted in the conceptual report:

1) Is there a strong degree of jointness beween commodity and non-commodity outputs that cannot be altered, for example, by changes in farming practices and technologies or by pursuing lower cost non-agricultural provision of non-commodity outputs?

2) If so, is there some market failure associated with the non-commodity outputs?

3) If so, have non-governmental options (such as market creation or voluntary provision) been explored as the most efficient strategy?

In a second phase, an OECD Workshop, held early July 2001, gives an opportunity to apply the analytical framework in an empirical manner, thereby indicating the sort of statistical analysis which is relevant to illustrate the "problematique" of multifunctionality. Some of the existing analysis, for example on agri-environmental indicators and on rural amenities, would, of course, be relevant, but it will be necessary to integrate them into a systematic framework. This shows that the policy requirements continuously create new challenges for statisticians, economists and policy analysts and advisers, which can only be addressed successfully if they work together in a dynamic context.

6. Conclusions

As stated by Berkeley Hill, two main factors lead to the need to reconsider the content of agricultural statistics: first, the changes in the reality of the agro-food sector, and in particular the structural changes at the farm level; and second, the evolving policy requirements, which reflect the changing expectations of society from the agro-food sector. These two factors are in any case closely related to each other, and create general pressure for greater flexibility in the system of agricultural statistics. I would like to emphasise in particular two facets of this necessary flexibility:

i) To create more synergies between agricultural statistics and other general economic and social statistics. This goal results first of all from the greater linkages between agricultural and non-agricultural activities (see the discussion of farm households in both papers by Berkeley Hill and Susan Offutt). From a policy viewpoint, it also reflects the fact that the agro-food sector will be increasingly influenced by non-agricultural policies, such as broad economic, social, environment, or regional policies.

ii) To create an institutional framework, within which the various evolving demands for "policy-relevant" (or more generally, "society-relevant") statistics are constantly confronted with the need to maintain the intrinsic quality and reliability of statistics. This means that statisticians (both official and academics) have to be in constant dialogue with policy advisers and policy-makers, and with all stakeholders, in particular farmers, consumers, representatives of the agro-food industry and environmentalists. The search for new forms of dialogue with the so-called "civil society" also applies to statistics, to the extent that the objective is to make them more relevant.

References

OECD (2001) Agricultural Policies in OECD Countries, Monitoring and Evaluation 2001
OECD (2000) Agricultural Policies in Emerging and Transition Economies
OECD (2001) Agricultural Outlook 2001-2006
OECD (2001) The Uruguay Round Agreement on Agriculture - An Evaluation of its Implementation in OECD Countries
OECD (2001) Differences in Production, Trade and Income Effects Among Farm Support Measures - The Crop Sector (forthcoming)
OECD (2001) Environmental Indicators for Agriculture, Volume 3, Methods and Results
OECD (2001) Multifunctionality - Towards an Analytical Framework
OECD (2000) Valuing Rural Amenities

Developed Country Agricultures: Preparing Statistical Systems for the Policy Needs of the New Millennium

Berkeley Hill
University of London, Imperial College, Wye, ASHFORD, Kent, TN25 5AH, UK.
e-mail: b.hill@ic.ac.uk

Abstract: Concepts behind statistics have to be adequately developed if the quality of information based on them is to be high. Changes in the structure of agriculture in developed countries have undermined the conceptual foundations of some important agricultural statistics. Concepts have to be rethought, and a short agenda of actions to do this is put forward, including a reconsideration of the basic unit of economic observation. To avoid obsolescence, the statistical system has to become responsive to future needs; conditions for flexibility are outlined

Keywords: Statistics, obsolescence, farm business structure, agricultural households

1. Introduction

Official agricultural statistics are a costly commodity yet can be highly valuable. They are used to inform society of the nature of the real world and changes taking place in it. They are the basis for assessing problems that exist or are developing and for which policy intervention may be required - the political significance of agricultural problems in developed countries is disproportionately large. Statistics can also be a tool for triggering policy instruments and for monitoring their performance (Holt, 2001), though they alone are unlikely to be sufficient for evaluation. In the US the returns to careful decisions about data and information have been found to be extremely high (Bonnen, 1975).

Civil servants are largely responsible for the output of agricultural statistics and their quality. In OECD countries various administrative arrangements can be found. In some, such as the UK, systems of agricultural statistics, even economic ones, have been very largely under the control of the Ministry of Agriculture, Fisheries and Food. In others, the national statistical authority takes a more prominent role (as in the Netherlands). However, all arrangements involve agencies (institutions), and statisticians form part of a bureaucracy.

The discipline of political economy can bring insights into the ways that bureaucracies operate, and has been used in explaining the difficulty of reforming agricultural policy both in the EU and US (Moyer and Josling, 1990). Lessons from this approach may be applied to systems of official statistics.

Some of the more potent predictions about the *supply* of statistics are as follows:

- Bureaucracies and their individual staff members have their own interests and agendas, and these will overlay the supply of statistics and their ability to service policies.

- Organisations require procedures and routines in order to operate, and in larger organisations there may be more people to consult before a change can be made, tending to make change slower.
- Substantial vested interests may have been built up in existing concepts, ways of measurement, analysis and dissemination that will be threatened by change, a problem that Arrow characterises as human capital being made obsolete (Arrow, 1974).
- Change itself requires resources to bring about; without extra resources, a fall-off in output of existing statistics is likely, which may be resented by users.
- Hence, bureaucracies will tend to prefer the status quo rather than one of change.
- Where changes are contemplated by bureaucracies, they will tend to be incremental rather than radical in nature.
- Radical change is contemplated only in crisis conditions, when the continued existence of the institution itself may be under threat.

In addition, political economy can be applied to the *demand* side:

- Users of statistics are often politicians or members of other bureaucracies, so their demands for statistics are again overlain by sets of personal and institutional objectives that may obscure those of the underlying policy. In the case of politicians these self-interests relate to their retention of power. In the case of bureaucrats, they relate to the prestige and influence of their institutions and to the position of individuals within them.
- Users of statistics are not always able to specify the statistics they require or to judge statistical quality, as they may not possess adequate information about the possibilities and limitations, a characteristic that parallels that of imperfect knowledge in consumer theory. This implies that statisticians must exercise an independent role as designers of statistics to meet declared and inferred policy needs.
- Different users will often require different formulations of what appears, superficially, to be the same information. This may lead to satisficing behaviour – taking the first option that appears to satisfy a need – which may lead to poor policy decisions.
- Where users are directly involved in providing resources for statisticians, there is danger that the independence of the statistical system will be undermined, and the provision of statistics will reflect the political and bureaucratic concerns of the users.

These characteristics, though perhaps uncomfortable to contemplate, flow from the inevitable way in which the production of statistics has to be organised and should not be interpreted as criticisms of present staff. Nevertheless they must be borne in mind when considering how the agricultural statistical system can adapt to the conditions of the new millennium.

2. Quality in statistics – the role of relevance and conceptualisation

Writers on statistics typically identify many of the same features of "good" quality, though the terms used may vary. Accuracy, coherence, consistency, continuity, timeliness, accessibility and presentation, comparability over time and space are normally mentioned (Brackstone, 1999; de Vries, 1998; Elvers and Rosen, 1998; Holt and Jones, 1998). All these may be classed as "intrinsic" properties of statistics. However, there seems to be a dichotomy of views about "relevance". On the one hand, Elvers and Rosen (1998) exclude it, as being what they term a *subject-matter problem*, whereas Holt and Jones (1998) and Brackstone (1999) regard it as a factor in the determination of statistical quality. What is apparent is that it is different in nature from the "intrinsic" characteristics, in that it is dependent on the validity of the link between what decisionmakers need on which to base their choices and actions, and what statisticians actually measure.

Relevance of a given array of statistics can change because of two main reasons. The first is that the *needs of users change* as the problems of the agricultural and food sectors evolve under long-term trends (economic, technical, political etc.) and emerge from short-term shocks. Unless the statistical system responds to changing user needs, its products will become outmoded; statisticians will find themselves ready to answer questions with which fewer users are now concerned and yet will be incapable of servicing new policy aims. An example of declining relevance is statistics on national self-sufficiency in food production. Examples of newer policy-related needs include agri-environmental indicators and statistics for rural areas. The paper by Gerard Viatte (OECD) in this volume deals with new policy challenges to statisticians, so these will not be covered here.

The second threat to relevance arises from *changes in the reality* on which statistics attempt to provide information. This paper focuses on this second aspect. Before the data on which statistics are based can be collected, there are the crucial stages of "conceptualisation" and "operationalisation". "Conceptualisation" involves developing concepts that are "capable of portraying and reducing the nearly infinite complexity of the real world in a manner that can be grasped by the human mind" (Bonnen, 1975). As concepts cannot be measured directly, "operationalisation" involves defining variables that are as highly correlated as possible with the aspect of reality that is being examined. In the US, Bonnen has stressed the significance of adequate conceptualisation if the agricultural information system is to perform satisfactorily (Bonnen, 1975; Bonnen, 1977). In the UK this concern has been expressed in relation to national accounts; Holt and Jones point out that "It is rare for the concepts that we strive to measure to be driven by a well defined theoretical construct"(Holt and Jones, 1998). However, only if this first step is reliable can "operationalisation" be undertaken adequately, leading on to the stages of measurement, data analysis and the production of statistics; ".. no matter how well one manipulates the numbers, one may still be measuring the wrong thing" (Bonnen, 1975). "Conceptualisation" is the responsibility of both statisticians (who constitute a major part of the "data system") and of members of the "inquiry system", outsiders who are not involved in the routine of actual statistics production and who therefore can contribute a more detached view (consultants, academics etc.).

Conceptualisation is not easy even in static conditions. In the dynamic economic and technical environment of the second half of the 20th Century, the changing nature of agriculture has presented a moving target, opening a gap between the conceptual basis of existing statistics and reality. Such shortcomings in statistics can be more insidious than failure in the "intrinsic" characteristics (inaccuracy because of poor response rates etc.) because conceptual obsolescence is not readily quantified and because it usually creeps in gradually. The need to generate statistics on a regular basis may divert attention from any widening gap, while the protection of vested interests and human capital in existing concepts and systems of measurement will tend to marginalize any gaps that are allowed to surface.

3. Conceptual obsolescence in today's agricultural statistics

The conceptual bases of many of today's agricultural statistics belong to a period when the industry was very different in structure from the present. The notion of the "family farm" seems to lie behind many of them – a potent but poorly specified unit that combines economic and social functions. This model implies; ownership and management of the business by principals who are related by kinship or marriage; the family lives on the farm; family members provide capital to the business; the family provides physical labour to the farm; and ownership and control are passed between generations (Gasson and Errington, 1993). We might add the perceptions that (a) the family's available labour is largely or wholly utilised within the farm (b) that the farm is the sole or main source of its livelihood, so that non-agricultural income can be ignored, and (c) management is typically hierarchical, with the head of the household being ultimately responsible and therefore justifying the title of "the farmer". The influence of this view of farming is easily seen in the main economic agricultural indicators used in the EU; these have methodological roots taken over from national systems that were established in the 1930s or earlier.

In the US, where a similar underlying assumption about the family structure of farming permeated the statistical system, the problem of obsolescence was recognised long ago, with major public debates taking place between statisticians and economists in the early 1970s and sporadically later (AAEA, 1972; Carlin and Handy, 1974; Baum, 1986; Bonnen, 1975; Bonnen, 1977; Bonnen, 1988; Bonnen, 1989). A parallel consideration took place in Canada (for example, see Loyns *et al*., 1983; Loyns *et al*., 1986). At the time, debate in Western Europe about statistical obsolescence was far less prominent, possibly because structural changes were less advanced here. However, in the EU the gap between concept and reality has now become much larger, and developments are in hand that will stretch the relationship beyond credibility. Some of the more prominent examples of conceptual obsolescence are considered below, though the first is particularly important as it affects those that follow.

3.1 The agricultural holding as an obsolete basic economic unit of observation.

For the purpose of compiling information on total numbers of livestock, crop output, land use patterns, labour input etc. the basic unit of observation is of only secondary concern; what matters is the quality of the aggregate. However, for purposes of economic analysis the integrity of the basic unit is critical. It has to be recognised that the production of agricultural commodities in OECD countries is frequently carried on by firms in combination with other activities. About a third of the people regarded as EU "holders" have other gainful activities, and probably a substantially larger share of farming couples. Though associated primarily with the smaller holdings, elements of pluriactivity can be found throughout the size spectrum, with the suggestion of a greater incidence as the largest farm sizes are reached. Households that run their farms as unincorporated businesses do not draw any impermeable boundary between their agricultural and other gainful activities, or between their functions as units of production and of consumption. Hence explanations of farming behaviour (pattern and intensity of land use, farm viability, investment levels etc.) cannot be taken in isolation. The overall situation of the complete unit (household-firm) needs to be taken into consideration. An extreme example of this is capital balance sheets, where practical and theoretical issues combine to make a balance that purports to relate only to farming near to meaningless (Hill, 2000a).

In creating the fictitious units of the "farm businesses" or "holding" (in farm accounts surveys such as the EU's Farm Accountancy Data Network – FADN/RICA) or the agricultural Local Kind of Activity Unit of the EU's Economic Accounts for Agriculture (EAA), the functional links within the real institutional unit (the unincorporated firm run by the households, or the incorporated businesses) are fractured. This dismembering is likely to produce observations that have limited use as data on which to base reliable statistics. As the US's AAEA Committee on Economic Statistics stated in 1972

> "Only when the basic economic structure of the industry can be described accurately by our data system will analytical accuracy be possible in dealing with the performance and behavioral characteristics that are the focus of most economic analyses".

3.2 Obsolescence of primary agriculture as a discrete economic activity

Another aspect of this first point is that the production of agricultural commodities (as classified in NACE.Rev1) is increasingly just part of a larger process. In the US attention has been drawn to the implications of vertical integration in the food chain, by merger or contractual arrangements, to the extent that for some commodities a farm gate price is no longer a meaningful concept (Vogel and Johnson, 2000). In the EU stress falls on the multifuctionality of agriculture and the combining of commodity production with the output of other goods and services; some may share resources with traditional farming (such as on-farm tourism, farm-based shops and farm forestry) or be dependent on the externalities of farming, while many may have only the capital base and management in common. In the UK all manner of gainful activity can be found combined with farming, a phenomenon that has a long history, appears to be growing and is undertaken for both economic, social and environmental reasons (Gasson and Errington, 1993; Harrison, 1975). Agricultural statistics have gone some way to recognise this situation; for example the Economic Accounts for Agriculture now allows the inclusion of "non-separable secondary activities" within the output of agricultural LKAUs (Eurostat, 1997a). However, the criterion of "separability" is rather arbitrary and users may assume that the EAA now covers all the activities of the institutional units, which it does not (Hill, 1998).

3.3 Obsolescence of the "one farmer per farm and one farm family per holding" assumption

Many agricultural statistics assume a hierarchical family structure that is increasingly unreal. Surveys frequently find that entrepreneurial decisions are taken by more than the senior person nominated as the "holder" or "head" of household, so that to take the characteristics of that person (age, education etc.) as the basis of an attempt to explain business behaviour would be misleading. Formal sharing of responsibility is intrinsic with legal partnerships but is often encountered as part of the normal process of intergenerational transfer in family businesses; on larger farms several families living in separate households may be involved. Such spreading of the entrepreneurial role is particularly problematic among businesses that have their own legal status; on company farms ownership and management are separated, and there may be no simple locus of responsibility for decisions. The person who takes short-term decisions may or may not be a director or shareholder. While many family-owned farm businesses are so structured solely for reasons of tax-reduction, and behave as *quasi-unincorporated firms,* this is not true of all the large-scale businesses that tend to dominate the volume of total output. In particular, it does not apply to the big units found in parts of Germany and in EU Candidate Countries. To attribute all residual income generated by these units to a sole fictional "farm family" is likely to be very misleading. Corporate farms and co-operative units also challenge the conventional basis of who are the intended beneficiaries of agricultural support policy.

3.4 Obsolescence of the view that the agricultural community comprises all farm operator families

Agricultural producers display a wide heterogeneity. Different statistical problems require appropriate concepts of the agricultural community to be applied. If the focus is on production (or land use), then all significant producers (or land users) must be included; this is the traditional stance of agricultural statistics. In one sense, all of us with kitchen gardens are producers, but our output is neither very relevant to our living standards nor significant quantitatively to national output. Thus in the EAA output from gardens can be excluded when applying the methodology to the EUR15 (Eurostat, 1997a). This is not valid for countries in Central and Eastern Europe where private plots generate a substantial volume of the national food supply.

However, a good case can be made that much of the focus of the EU's Common Agricultural Policy (and of many national policies) is primarily social in nature – supporting the standards of living of the agricultural community. Despite its prominence in statements of EU policy aims, the "agricultural community" is a poorly defined concept (Hill, 1990). As a social target group, it seems reasonable to include only those households that are primarily dependent on farming. When developing its Income of the Agricultural Households Sector (IAHS) statistics, Eurostat has covered only households where farming is the main income source (initially of the whole household, now of the household reference person)(Eurostat, 1995). Judged on the basis of countries where data exist, only perhaps half of all EU households with some self-employment income from farming qualify for inclusion. On many, predominantly small, holding there will be no agricultural household, whereas on other, mainly larger ones, there may be several.

At an early stage of shaping the IAHS methodology the senior Agricultural Statistics Committee advised that households headed by waged (or salaried) farm workers should not qualify as "agricultural". This decision has presented difficulties for family farms arranged as companies, as directors and managers are in law employees of their own company; technical fixes are used to get round this problem. The issue has re-emerged much more problematically with the enlargement of Germany and the prospective accession of countries that include large-scale units; families working on them are clearly seen as potential beneficiaries of agricultural policy, and to exclude them would distort the picture emerging from IAHS statistics. Thus the concept of the agricultural community is having to be rethought in the light of historical developments.

4. A short agenda for improvement

As noted above, this paper is primarily concerned with the implications for statistics of changes in the nature of agriculture itself rather than with new policy directions. Among the latter there is an obvious need for better information on non-farm businesses in rural areas, regional and local-area statistics, the development of agri-environmental indicators that add an economic dimension to physical data, supply chains (including animal movements) and so on. But a list of improvements emerges even from the issues covered in this paper. These include the following:

4.1 Determine the basic statistical unit.
While the "holding" may be adequate for assembling aggregate physical statistics, it is not satisfactory as the basis for economic statistics; this requires the use of a real institutional unit. But the precise nature of this unit has to be established, such as the definition of what constitutes a single business and, among unincorporated farms, the borders of the household that operates it.

4.2 Develop a flexible typology of real institutional units that reflects both legal structure and policy needs.
Legal status is already relatively well established (of which the main classes would be unincorporated, incorporated and other). A further step needs to be taken to differentiate within classes according to policy need. Dependence on farming for the main income source is an obvious first step (which would enable a separation of agricultural units from non-agricultural ones) though a further differentiation along the lines used in the US might be considered for wider application (limited resource, retirement, residential/lifestyle, primary occupation family farms according to economic size, nonfamily farms etc.)(USDA,1998)

4.3 Develop accounts based on real institutional units.
Accounts based on real agricultural institutional units have advantages of easy interpretation, avoidance of arbitrary allocations associated with creating fictional units, and meaningful balance sheets. For the agricultural households sector, such accounts provide the opportunity to measure disposable income, an indicator more appropriate to the living standards aim of agricultural policy than those currently in use. Such an approach is the FAO preferred basis for aggregate economic accounting for agriculture (FAO, 1996). Eurostat's IAHS statistics are a step in this direction. A fuller schema of such accounts for European statistics has been proposed to set alongside the traditional activity-based accounts, which could then be simplified (Hill, 2000b).

4.4 Integrate aggregate and microeconomic accounts.
Issues in agricultural policy often require distributional information (for example, levels of income by farm size, type and region or by the socio-economic characteristics of their operators). For historical reasons, the EU-level systems of accounting at aggregate

and microeconomic levels developed separately and have not been well integrated, causing problems among users. A potential advantage of basing both levels of account on the same real unit is an improvement in this situation.

4.5 Adapt data collection systems.

Developing new concepts and making them operational is of little use unless measurement is possible to generate new data that can be turned into information. Making changes to data collection can be costly and is likely to face practical difficulties; it is at this stage that vested interests and bureaucratic inertia can be most potent. However, shifting attitudes among farm operators and politicians (for example, a greater willingness to discuss non-farm incomes), coupled with new information technology, may offer opportunities for change in the new millennium that were not possible even five years ago.

5. Preparing statistical systems for change

Situations have been identified above in which conceptual obsolescence has been allowed to permeate the statistical system because of a failure to adapt to changing conditions in the agricultural industry. Changing policy objectives could have provided another set of examples. Today's structure and policy aims require a statistical system that is based on concepts different from those that were appropriate when the methodological foundations of the present statistics were established. However, to simply re-engineer agricultural statistics on a new set of concepts appropriate to the new millennium would be to miss a major point – that statistics have to be constantly updated. As was pointed out in the US almost thirty years ago, "Ideally, what we need is not a *right* definition of a concept or even a *right* concept, but a system in which we have the flexibility to match concept and measurement to differing and changing objectives" (AAEA, 1972). The point has been restated for the EU by Holt (2001).

How may this flexibility be achieved? Three features seem to be necessary. *Firstly*, there must be regular and frequent tripartite discussions between users of statistics, providers of statistics and members of the "inquiry system", such as independent academics and consultants. It appears that users and providers by themselves are not capable of maintaining sufficient objectivity to counter conceptual obsolescence without the added stimulus of the third group. Academics should be able to provide contextual insights, such as the long-term economic and political trends and (perhaps) the risks of shocks. Their role in statistics is rather similar to that of the independent evaluator within policy assessment; while those closely involved in the administration of policy have detailed knowledge of the way it works and are in the best position to provide answers, they cannot be relied on to ask all the right questions. Examples exist of attempts to bring these three groups together, such as in Canada by a series of professional advisory committees and consultations (Fellegi, 1996, quoted in Brackstone, 1999). However, in the EU attendance at such meetings tends to be dominated by suppliers (Eurostat, 1997b).

The *second* necessary condition seems to be that of instilling a culture of change among the suppliers of statistics. Policy analysts often note the tendency of bureaucracies that administer policy to prefer established procedures and mechanisms even when old problems have disappeared and new ones emerged. Policy succession and termination is difficult to achieve unless it is built into the system by, for example, only providing funding for a finite period, after which action would cease unless a continuing need could be demonstrated (Hogwood and Gunn, 1984). Parallels in statistical bureaucracies are to be expected, and perhaps the solutions are similar.

The *third* necessary condition is that adequate resources must be devoted to the change process. When managers of statistical systems are faced by budgetary constraints, there will be a tendency to give priority to current operations and squeeze expenditure on activities of a strategic nature, carrying the danger of conceptual obsolescence and loss of relevance. Here is a classic case of a conflict between what is urgent and what is important.

The problem of keeping agricultural statistics up to date is by no means new. At the start of the new millennium statistical systems in developed countries have some catching-up to do and are required to become more responsive to changing conditions. Both activities are resource-demanding, at least in the short-run. However, the cost to society of policies based on poor information is likely to be far greater.

References

AAEA (1972). Our obsolete data systems: new directions and opportunities (American Statistical Association - American Agricultural Economics Association Joint Committee on Agricultural Statistics). *American Journal of Agricultural Economics*. **54**, 867-80.

Arrow, K. J. (1974). Organisation and Information. *In* "The Limits of Organization", pp. 33-43. W. W. Norton, New York.

Bonnen, J. T. (1975). Improving Information on Agriculture and Rural Life. *American Journal of Agricultural Economics* **57**, 735-63.

Bonnen, J. T. (1977). Assessment of the Current Agricultural Data Base: an Information System Approach. *In* " A Survey of Agricultural Economics Literature - Vol. 2. Quantitative Methods in Agricultural Economics, 1940s to 1970s.", (L. R. Martin, ed.). University of Minnesota Press, Minneapolis.

Brackstone, G. (1999). "Managing Data Quality in a Statistical Agency," Statistics Canada, Survey Methodology, Catalogue No. 12-001-XPB, Statistics Canada, Ottawa.

Carlin, T. A. and Handy, C. R. (1974). Concepts of the Agricultural Economy and Economic Accounting. *American Journal of Agricultural Economics* **1974**, 964-75.

de Vries, W. F. M. (1998). How are we doing? Performance indicators for national statistical systems. *Netherlands Official Statistics* **13**, 5-13.

Elvers, E. and Rosen, B. (1998). Quality concepts for official statistics. *In* "Encyclopaedia of Statistical Sciences, update volume 3", pp. 621-29. Wiley-Interscience, New York.

Eurostat (1995). "Manual on the Total Income of Agricultural Households (Rev.1). Theme 5 Series E. Eurostat, Luxembourg.

Eurostat (1997a). "Manual on the Economic Accounts for Agriculture and Forestry (Rev.1)," Eurostat, Luxembourg.

Eurostat (1997b). "Towards a new generation of agricultural statistics: user needs Symposium. Draft report on Symposium of 27 October 1997.," Eurostat, Luxembourg.

FAO (1996). “A System of Economic Accounts for Food and Agriculture. FAO Statistical Development Series 8,” Food and Agriculture Organisation of the United Nations, Rome.

Fellegi, I. P. (1996). Characteristics of an Effective Statistical System. *International Statistical Review* **64**.

Gasson, R., and Errington, A. (1993). “The Farm Family Business,” CAB International, Wallingford.

Harrison, A. (1975). “Farmers and Farm Businesses in England,” Misc. Studies 62, Department of Agricultural Economics and Management, University of Reading.

Hill, B. (1990). In search of the Common Agricultural Policy's "agricultural community. *Journal of Agricultural Economics* **41**, 316-26.

Hill, B. (1998). The Implications for Agricultural Statistics of Changes in the System of National Accounts. *Journal of Agricultural Economics* **49**, 359-77.

Hill, B. (2000a). The illusory nature of balance sheets in agricultural economic statistics: a note. *Journal of Agricultural Economics* **51**, 463-467.

Hill, B. (2000b). Linking the established subsystems of economic statistics to provide a rounded view of agriculture. In: Pauli, P, (ed.). *In* “Agricultural Economic Statistics, Proceedings of the Seventh IWAG-AGRI Seminar, Luxembourg 5-7 July, 2000”, (P. Pauli, ed.), pp. 46-74. Eurostat, Luxembourg.

Hogwood, B. W. and Gunn, L. A. (1984). “Policy Analysis for the Real World,” Oxford University Press, Oxford.

Holt, T. (2001). Official Statistics and their Contribution to Public Policy. In: *Information and Knowledge: The role of statistics.* Proceedings of the 86th DGINS Conference, Porto, June 2000. Theme 1 General Statistics. Eurostat, Luxembourg.

Holt, T. and Jones., T. (1998). Quality work and conflicting policy objectives. *In* “84th DGINS conference, 28-29 May”. Office of National Statistics, Stockholm.

Loyns, R. M. A., Freshwater, D., and Beelan, G., eds. (1983). “Proceedings of the Seminar on Farm Income Statistics”, Department of Agricultural Economics and Farm Management, University of Manitoba,, Winnipeg.

Loyns, R. M. A., Freshwater, D., and Hamilton, E. F., eds. (1986). “Proceedings of the seminar on the theory and practice of agricultural wealth accounts. Occasional Series No 16”, Department of Agricultural Economics and Farm Management, University of Manitoba., Winnipeg.

Moyer, H. W. and. Josling, T. E. (1990). “Agricultural Policy Reform: Politics and Process in the EC and USA.,” IOWA State University Press, Ames, Iowa.

USDA (1998). "Structural and Financial Characteristics of U.S. Farms 1995: 20th Annual Family Farm Report to the Congress". Resource Economics Division, Economic Research Service, USDA, Washington.

Vogel, F. and Johnson, J (2000). Measuring agricultural income with the changing face of agriculture. *In* “Proceedings of the Seventh IWG-AGRI seminar”, (P. Pauli, ed.), pp. 257-284. Eurostat, Luxembourg.

5 June 2001

Invited Paper Session

EVOLUTION OF THE ROLE OF AGRICULTURE IN DEVELOPING COUNTRY ECONOMIES

Chair: J. Bruinsma

The Evolution of Agriculture's Role in Economic Development: the Case of sub-Saharan Africa

Hans Binswanger
Director, Environmental, Rural and Social Development Department, Africa Region, The World Bank
e-mail: HBinswanger@Worldbank.org

Abstract: Both the socialist and capitalist traditions of development economics accepted the presumption that industry led and agriculture lagged. The intellectual demise of this approach opened the way for a reexamination of agriculture's contribution to growth and development. The case for agriculture as an engine for economic growth was enhanced through linkages that growth in the agricultural sector has with income, and hence, rural demand for consumer goods and services from the non-farm economy. African agricultural growth has been slow for many reasons. A combination of measures are needed to increase it. There is an urgent need for building statistical capacity in Africa in support of poverty-reducing agricultural growth.

Keywords: Economic growth; forward, backward and consumer demand linkages; African agricultural growth; integrated statistical capacity building

1. Why focus on agriculture and why now ?

In Africa, many countries will become eligible for debt relief under the Highly Indebted Poor Countries (HIPC) Debt Initiative. In that context they have been asked to prepare Poverty Reduction Strategy Papers (PRSPs), which in turn are creating significant demands for relevant data for monitoring progress on poverty reduction. PRSP processes are participatory, which subsequently generates increased demand for data from social society.

Given the concentration of 70 percent of the poor in rural areas, agriculture must play a central part in poverty reduction. But sub-Saharan Africa's growth has been slow. Our understanding of agriculture's contribution to economic growth has changed over time, and so has the nature of the contribution. Equally important is our ability to articulate this topic persuasively to influence policy and programs.

Market failures and government failures are known to depress growth and leave a legacy of poverty. The impact of intellectual failures and failures to communicate may be less obvious, but equally damaging. The economics profession as a whole, and development economics in particular, suffered such a failure in the period immediately after World War II, when a development paradigm triumphed that justified the plunder of agriculture in the interests of growth based on protected and noncompetitive industry. Poor people in rural areas throughout the world continue to live with the material and social consequences of that mistake. Although the lessons of that particular past error have in general been internalized, we may not yet have an intellectual consensus within the development community on how to reflect these lessons in policy and in the design of programs of assistance.

Much of the recent controversy about the relative importance of growth and redistribution arises because the discussion takes place at the macro level; e.g. does more rapid aggregate economic growth necessarily translate into lower poverty? When the discussion is taken down to the level of a village in which 90 percent of the people are poor and rely on agriculture for their primary livelihoods, it becomes clear that in most cases significant growth in agriculture in those very villages will be a necessary step toward improved lives. It will, however, not be the only answer. Some people will choose to migrate, others will find off-farm employment, and others will be excluded from any improvement at all, and will need social assistance. Yet a viable strategy of poverty reduction in rural areas, particularly in Africa, will have to be predicated on income growth derived from agriculture.

Many of the countries qualifying for debt relief are in Africa, and much of the poverty is rural. Unless a clear understanding of the role of widely shared agricultural growth and how to achieve it is put squarely on the table during the process of formulating poverty reduction strategies and monitoring expenditures to implement them, we run the risk of participating in a second missed chance for rural development. Quite simply, the renewed opportunity could by-pass agriculture, and encourage attention to improved primary education, better health services, clean water, better roads: all important and appropriate investments, but not sufficient in and of themselves to generate increased earnings from agriculture.

We must listen to rural people. They consistently say that in order to become less poor, they need to raise their incomes derived from agriculture. They are right, and this paper attempts to provide insights from the economic literature to explain why that is so.

2. An evolving understanding of the role of industry and agriculture

In both the socialist and capitalist traditions of development economics, academics accepted the presumption that industry led and agriculture lagged (Schiff and Valdes, 1998). The resulting pre-eminence accorded to industrial policy had little or no empirical foundation, and represented a monumental failure on the part of development economists. The lack of logic in this course was characterized succinctly by Schultz (1964) when he wrote that "economists who have been studying growth have, with few exceptions, put agriculture aside in order to concentrate on industry, despite the fact that every country has an agriculture sector and in low-income countries it is generally the largest sector."

Lack of confidence in agriculture as an engine of growth led to a strategy based on import substitution, since the countries adopting the strategy were net agricultural exporters and net importers of industrial and finished goods. Protection of industry followed naturally, since the industries expected to lead national growth were not, for the most part, internationally competitive. Thus lack of confidence that agriculture could contribute to growth led to adoption of a set of policies assuring that it could not. Among these were tariff barriers, differential exchange rates for different sectors, over-valued exchange rates penalizing exports, implicit taxes on agriculture levied through monopolistic marketing boards, subsidized lending for favored industrial schemes, and other measures. The distortions and transfers of resources to favored industries were justified as temporary measures to be dismantled once the infants had matured and were able to compete. The costs of this protection were felt widely across entire economies, with disastrous effects on growth.

The intellectual demise of import substitution opened the way for a reexamination of agriculture's contribution to growth and development. The writings of Johnston and Mellor (1961) and Mellor (1966) encouraged economists to view agriculture as a positive force in economic growth and development. These authors also stimulated debate on the interdependence of agricultural and industrial growth which, in turn, led to a growing interest in the measurement of intersectoral resource transfers. The linkages between agriculture and the rest of the economy assumed under the structuralist view to be unimportant were reexamined.

Many early studies of linkages between agriculture and broader growth focused largely on production linkages arising from the establishment of new productive activity (Thirtle *et al.*, 2000). These linkages were classified into backward and forward linkages. Backward linkages consist mainly of derived demand for inputs from the new activity. Forward linkages consist of the induced creation of new productive activities from placement of a new intermediate product on the market.

The case for agriculture as an engine for economic growth was enhanced through increased focus on the impact that growth in the agricultural sector has on incomes, and hence, demand for consumer goods and services from the non-farm economy. Against the background of India's experience with the green revolution, Mellor (1966, 1976) pointed out that although production linkages from the agricultural sector may in fact be weak, especially in the case of subsistence agriculture, consumption linkages from the agricultural sector do have major indirect effects on the rest of the economy. According to Mellor's view, the impact of agricultural growth on other sectors in the economy depends on how much of the extra wage income is spent on non-agricultural goods and services, leading to a new round of spending, and how much is spent on the increased food production itself, or leaks out into savings. Later researchers clarified that the linkages are limited to the extra spending on tradable goods and services, since imported goods are produced in other countries.

The 'linkages paradigm' (after Delgado *et al.*, 1994) was applied specifically to rural areas in the 1960s and 1970s in response to the writings referred to earlier, but also in response to the concern over rapid urbanisation. The flow of rural migrants to cities spurred unemployment because the more capital intensive urban economy could not provide jobs for the growing population. Economists and policy makers perceived a need to create jobs outside agriculture and outside cities. This, according to Delgado *et al.* (1994), led to a focus on growth processes boosting demand for rural non-farm activities. The studies carried out during the 1970s and early 1980s, found that consumer demand was quantitatively the most important linkage. What spurs rural non-agricultural growth is the demand for "rural non-tradables" such as housing and other goods and services which must be produced close to where they are consumed.

The explosion of the rural non-farm economy in China after the reforms of 1978 and less spectacular but still substantial growth in South Asia and parts of Latin America confirmed the observation that where rural incomes are rising, much of the added wealth is due to non-farm rural employment. Unlike the earlier structuralists, researchers of the 1980s did not confuse cause and effect. The interest in rural non-farm employment did, however, spur an effort to discover mechanisms to stimulate employment and increase activity in this sub-sector. The effort has led to further elaboration of the linkage theories, and hence directly back to farms, since growth in agriculture provides the primary stimulus for growth off the farm in rural areas.

Agriculture need not be the only source of incremental income and consumer demand in villages and small towns. In some places tourism fulfils the same function, especially labor intensive tourism employing local people. Mining, forestry, fishing, and other natural resource based activities are important in some localities. Any activity causing a tradable good or service to be sold out of the community and generating income can stimulate the requisite multiplier effects through linkages. At present and throughout most of sub-Saharan Africa, agriculture is clearly the most likely candidate.

3. The performance of African agriculture

Given the key role that agriculture must play in many African countries, the poor performance of the sector is doubly troubling. Growth is foregone in agriculture and jobs and incomes are foregone in rural non-farm activities.

The historical view of overall global economic growth provided by Maddison (1996) presents striking evidence of Africa's poor performance. Maddison's cross-regional study suggests that African real GDP grew at 0.6 percent per annum between 1820 and 1992, which was half the rate of world growth. Since much of GDP is still in agriculture, the sector has done poorly. African agriculture lags other regions both in production and productivity. For the two decades prior to 1990, agricultural production per capita declined significantly. Cereal yields, on average, are less than half those in other developing regions. Even for tubers and plantain, for which African agro-ecological conditions are suitable, average yields are lower than in Asia and Latin America.

The sector is severely under-capitalized, and agricultural labor productivity is correspondingly low. African agriculture has not accumulated capital, and has not grown fast enough to realize its potential as an important engine of economic growth and poverty reduction.

The adverse performance of African agriculture can be attributed largely to poor policies and institutional failures that have plagued the continent for several centuries. But adverse endowments may also have served to lock in poor policies, and to increase the difficulties of breaking out of the poor policy trap. A modest acceleration of agricultural growth during the 1990s offers some grounds for optimism. Such growth must occur if the ambitious targets for poverty reduction are to have a chance to succeed.

3.1 Why has African agricultural growth been so slow?

Explanations for slow growth are many, but they generally fall into one of two categories: adverse resource endowments and failures of policies and institutions.

Adverse resource endowments. Hayami and Platteau (1997) argue that Africa's resource endowment, in particular its abundance of land and low population density, have retarded agricultural growth. It is ironic that land abundance, the corollary of low population density, should be seen as a curse, rather than a positive factor in agricultural growth, since in much of the world land hunger is a contributor to rural poverty. Low population density increases transportation and transactions costs. Sparse settlement patterns impede emergence of active competitive markets in outputs and inputs, as low volumes discourage entry of traders in sufficient numbers to sustain a competitive environment. The costs of providing agricultural and social services are high. Rural financial markets struggle in concert with product and input markets. Demand for credit is low, and supply is constrained by covariance of income, the lack of diversification, and the absence of a suitable collateral, as land has little value. New technologies enter slowly or bypass such areas, as high transaction costs and capital shortages impede adoption. Hayami and Platteau argue that adverse resource endowments bear a primary responsibility for the failures of agricultural and rural development in Africa.

Other studies have focused on other adverse conditions hindering agricultural development in Africa. Many countries are landlocked, which further increases transactions costs (Bloom and Sachs, 1998; Collier and Gunning, 1997). Land quality is poor (Voortman *et al.,* 1998; Donovan and Casey, 1998). Risks of drought are high on 60 percent of African land and almost half of the land is unsuitable for rainfed cultivation because the growing period is too short (UNCTAD, 1998). Finally, endemic livestock diseases (Coetzer *et al.,* 1994) and human diseases (Bloom and Sachs, 1998) exact a devastating toll, the most destructive of which are malaria, tuberculosis, and more recently the HIV/AIDS epidemic (UNAIDS, 1998; McCarthy *et al.,* 2000).

Centuries of poor policies and institutional failures. Poor policies and adverse institutional frameworks are not a recent phenomenon in Africa, but have prevailed with limited interruptions for several centuries.

Pre-colonial. Extraction from rural areas of Africa during the pre-colonial era occurred through the depredations that the slave trade exacted on economic, political and social life. Especially between 1650 and 1850, trade in human beings disrupted Africa's natural demographic development as well as its social, institutional, and moral development (Fage, 1977; Aplers, 1977; Curto, 1992). Slavery prevented the emergence of well performing states not only in the areas raided, but even in the slave trading states. Meillassoux (1981) showed that the political entities conducting the slave raids did not accumulate capital or a productive base, and were never able to reproduce themselves. They even failed to reproduce the population of already captured slaves. African rural people became a non-renewal resource extracted from an ever widening radius of destruction of the economic, social, and political foundation supporting subsistence agriculture.

Colonial. With the abolition of the slave trade and the onset of the colonial era, extraction from rural areas did not end, but shifted from people to their labor and the products of it. Several policy and institutional mechanisms were developed to ensure the capture of the gains from trade by the colonial elites, with restricted market access for indigenous populations. In many East and Southern African countries peasant farmers were prohibited from producing and selling cash crops. In other countries they had to sell to monopolistic companies at depressed prices. Differential taxation and distortions were widely used to induce peasant farmers to supply labor to plantations, to settler farmers, to mining sectors, and for public works (Binswanger *et al.,* 1995).

Post-colonial. A virulent combination of nationalism,agricultural pessimism, and centralization overtook many African leaders in the post-colonial period. As a consequence, opportunities to foster agricultural development and allow rural areas to recover from the ravages of the colonial and pre-colonial eras were foregone. Agriculture was heavily taxed and farm profits and welfare depressed, largely through producer prices held below world price equivalents by administrative means.

The negative effects of heavy taxation and extraction on agricultural growth could have been partially mitigated if the ruling elites had set taxes low enough and invested some of the surplus in rural public services and infrastructure. East Asia provides some twentieth century examples of relatively high extraction from rural areas coupled with substantial public investment in small-holder agricultural and rural development (Karshenas, 1998). In Africa rural public investment has typically been low, and subsidies for fertilizer and credit have usually benefited large farmers and other members of the rural elite, rather than small producers.

Sachs and Warner (1998) find that poor economic policies of the post-colonial era have played an especially important role in slowing growth. Their econometric analysis suggests that despite the climatic, geographic and demographic trends, all of which were estimated to hinder growth, Africa could have achieved per capita growth of 4.3 percent per annum, as opposed to 0.8 percent, if policies more conducive to growth had been adopted.

Adverse post-colonial policies included urban bias which remains in many African countries, and the agricultural interest groups able to influence policy are often the rent seekers in the marketing sector whose success comes at the direct expense of poor producers. Africa's enduring poverty, predominantly rural populations, closed political systems, and weak institutions of civil society all contribute to the persistence of urban bias (Lipton, 1993).

Post-colonial regimes established highly centralized political, fiscal and institutional systems for rural development. These high levels of centralization inhibited the development of local institutional capacity, limited local resource mobilization, undermined accountability of development programs to local populations, and inhibited their participation (McLean *et al.*, 1998). Lack of democracy in most of the countries and suppression of voluntary private associations further inhibited local initiative and participation.

Adverse trade regimes of OECD countries. Africans were not alone in penalizing their agricultural sectors. Developed countries joined in during the second half of the twentieth century by erecting barriers to the products that Africa could export competitively. Protectionist measures of the OECD countries have clearly accelerated the declining trend in world agricultural prices and therefore limited the export and the growth potential of agriculture in the developing world, including Africa (Binswanger and Lutz, 2000). Subsidies to agriculture in members of OECD equal Africa's GDP. These transfers are largest in the European Union, with Japan and the United States transferring income at just over half the EU level (Josling, 1998). Several developing countries (Brazil, Thailand) have managed to penetrate developed country markets for some products despite such restrictions. But Africa, for the most part has not. The trade barriers are self-reinforcing, since the resulting depressed prices and poor export prospects inhibit investments in productivity and quality that would make African products more competitive.

Inhibiting effects of poor policies. This brief tour of history suggests that over the past several centuries, opportunities for private individuals or groups to engage in free, competitive trade and investment in African agriculture and agro-industry were extremely limited. Incentives for farmers to put cash and labor into their farms and their natural resource base were minimal. Active public investment for agriculture and rural development rarely continued for extended periods, and even when in place was over-centralized and dominated by inefficient public entities. Little of the tax collected in rural areas returned to improve public services and infrastructure. Communities could not use local tax bases for their own development, as economic activity was depressed by the heavy level of agricultural taxation. Moreover, all significant local tax bases were assigned, by design or default, to colonial or central governments, or to monopolistic private or state structures.

The role of conflict. Conflict has been a consistent scourge of the African continent, exacerbating the impact of poor policies. The effect of conflict on agricultural performance can be observed in many countries. Sudan's considerable agricultural potential has been locked out by its 16 year civil war costing the lives of about 2.5 million people. The country has the largest number of internally displaced people in the world, hardly an environment conducive to sustained agricultural growth. Conflict is chronic in Angola, Burundi, Democratic Republic of Congo, and Somalia, and has recently flared up in Guinea-Bissau, Liberia, and Sierra Leone. Mozambique and its recent rapid growth demonstrate how agriculture benefits when conflict ends.

A growing literature suggests that low population density, remoteness from markets, and abundance of natural resources on the one hand, and conflict and adverse policies and institutions on the other, are indirectly linked (North, 1989; Tilly, 1990; Rueschemeyer *et al*., 1992; Collier and Binswanger, 1999; Goreux, 2001). The adverse impact of endowments can take additional forms, even when the endowment appears to be favorable. Collier and Binswanger explain how high mineral wealth can also lead to poor policies and to conflict.

The Lack of Voice of Rural People. Dispersed rural populations face insurmountable difficulties organizing themselves to have political voice. That the majority of African farmers are women has made political organization of producers even more difficult. The persistence of adverse policies reflects the much greater capacity of the rural and especially the urban elite to organize themselves relative to small farmers. The elite are therefore able to control policies, institutions, and the distribution of public resources. As noted, lack of open political systems and of well-articulated, competitive institutions in civil society have characterized these systems.

3.2 Recent policy and institutional changes and their impact

In recent years sentiment in Africa in favor of democratic governance has risen, and macroeconomic and agricultural reforms have begun to improve the competitiveness of agriculture. The exchange rate overvaluations of the 1970s and 1980s have been largely corrected, and inflation and budget deficits are down. Trade policies still raise import prices, but the anti-export bias has eased. But while macroeconomic stability has generally improved, in many countries it remains fragile. Financial sectors need strengthening. And institutions and rules are weaker than in other regions, diminishing investor confidence (Brunetti *et al*., 1997).

The resumption of agricultural growth in a number of countries since 1990 is in part in response to improved policies and institutions. Between 1990 and 1997 in twelve countries agricultural growth rates exceeded 4 percent annually (Benin, Cameroon, Chad, Guinea, Guinea-Bissau, Equatorial Guinea, Lesotho, Malawi, Mauritania, Mozambique, Namibia, Togo). After 1993 (1993-97) six more countries joined this group (Angola, Côte d'Ivoire, Ethiopia, Mali, South Africa, Zimbabwe). These trends represent a significant improvement from the 1980s when only 3 countries had annual growth rates exceeding 4 percent (Benin, Guinea-Bissau, Togo). Indeed, there is much evidence that African farmers and rural nonfarm entrepreneurs respond to incentives, both positive and negative.

Clearly, policies matter, even in an environment where there are serious material constraints. The fact that policies were poor for the entire century would lead one to believe that there are unrealized opportunities for further growth of agriculture. Providing the right policy environment for agriculture to grow is critically important in the African context, as agriculture remains the main source of rural livelihoods, accounting for about 35 percent of the region's GDP and 40 percent of exports.

4. Actions to spur agricultural growth and rural poverty reduction

A significant number of countries have not completed the policy and institutional reform agenda, and/or are experiencing second generation problems associated with the implementation of policy reforms. Private agents have not entered sufficiently into input, output, and rural financial markets, and the degree of market development and competition remains low. Tariff and nontariff barriers to agricultural and agro-industrial trade continue to be high inside Africa and in industrial countries. Public expenditure allocations for rural areas remain inadequate. Decentralization of public agricultural and rural development services is proceeding slowly in most countries, with fiscal decentralization still lagging badly. Thus a substantial agenda remains. In order to spur agricultural growth and rural poverty reduction, one must:

1. Further improve the policy environment and complete necessary reforms.

This includes: further stabilizing macroeconomic policies; enhancing the institutional framework, reduce corruption, and strengthen the credibility of rules; removing trade barriers in and among African countries; improving OECD market access; agricultural and agro-industrial growth cannot occur if producers are confined to local markets.

2. Improve domestic agricultural policies

This includes: dealing with remnants of Marketing Boards and other entities or hindrances restricting private sector entry; enhancing producer organizations and the private agri-business environment; improving input availability and markets; lowering transaction costs.

3. Accelerate political, administrative and fiscal decentralization

The excessive centralization of African states makes it very difficult to finance and implement development programs, and it dis-empowers rural and urban populations alike. Decentralization includes five main dimensions: empowering communities, empowering local governments, realigning the center, improving accountability, and building capacity.

4. Provide adequate public and private investments for rural areas

To accelerate agricultural growth and rural development, large investments will be required, both by the private and the public sector. Needed public investments, partly or fully financed by central and local governments include investments in education, health, transport infrastructure, water supply and irrigation, electrification and communication, and agricultural research and services. The last two decades have not seen an expansion of these investments; rather the succession of macro-economic crises and adjustment programs has cut deeply into their financing. Convincing governments to support these investments directly, and to create the decentralized institutional environment where local populations can partly undertake and finance them, remains an enormous challenge.

Much private investment is of course also needed. On-farm investments include investments in agricultural inputs, livestock, tree capital, soil improvements, irrigation, farm machinery, housing, and human capital. Agro-industrial investments are required into plant and equipment, as well as into skills, operating systems, and market development. Farmers and the private sector will have to drive this investment response.

5. Provide a voice to the rural poor

External pressure has been essential to accomplish the limited policy reforms already achieved. But imposing conditionality does not lead to sustained policy and institutional reforms. Instead, small-holder farmers, and other poor groups need help to organize and form an effective coalition with agro-industrial interests in the cities. They need to be able to enter into alliances with external partners to make them an effective force in the political bargaining process in which policies and institutions are determined, and sustained. Since the majority of African farmers are women, helping them acquire literacy and organizational abilities is of the highest priority. They must be empowered and become better informed about all aspects of rural development. The question for your profession is how you can best serve those information/empowerment needs. Establishing a multi-sectoral focal point in government serving the rural sector more broadly would also help.

6. Scale up efforts to address HIV/AIDS in rural areas

AIDS in Africa constitutes a massive threat to development, including threatening the gains made in agricultural growth. Estimates suggest that instead of life expectancy rising to 64 years by 2010-2015, a gain expected in the absence of AIDS, life expectancy will regress on average to 47 years, a loss of 17 years, and a level similar to those experienced in the 1950s (UNAIDS, 1998). While only one tenth of the world's population live in sub-Saharan Africa, 83 percent of all AIDS deaths since the start of the epidemic have been in the region, most of these being between the age of 15-49.

Apart from the enormous pain and suffering it inflicts, AIDS has now emerged as the greatest development challenge. In agriculture it undermines the incentives and the ability to save, and leads to disposal of assets under distress when households have to take care of AIDS victims and the orphans left behind. Since it hits mainly young adults, it sharply reduces the labor supply, and increases dependency rates, a direct determinant of poverty. Cropping patterns will shift in favor of less labor intensive crops and it will be more difficult to undertake the necessary labor investments in land and animal husbandry, and in housing and public infrastructure. Effective scaling up of efforts to address HIV/AIDS in rural areas is a key determinant to ensuring sustained gains in agricultural growth and development.

5. The need for an intensive effort of statistical capacity building in support of rural poverty reduction work

Some years ago I helped initiate a statistics initiative for Africa focusing on agriculture, food and natural resources. Unless we give attention to this, we will still be in the dark in years to come about the performance of agriculture, its linkages, and its poverty impacts. FAO and USDA have come on board and we have sought to jointly move the initiative forward. The African Commission for Agricultural Statistics endorsed the approach at its biannual meeting in Conakry in May 1999. It has generally been agreed that:

- It is essential for statisticians to better understand the demand side for data. This requires interacting with decision-makers and analysts who use statistics for various purposes.

- One must find cost-effective and accurate ways of providing just on time information in budget-constrained and institutionally weak circumstances. We need best practices adapted to countries' institutional capacities and circumstances.
- The production of agricultural statistics must not only be related to the demand side, it also much be fully integrated within broader efforts at statistical capacity building at the national level, and thus seek to serve the broader data needs for poverty reduction and growth including monitoring and evaluation.

To support these objectives, a proposal has been developed entitled: "Programme for Strengthening National Systems of Food and Agricultural Statistics in Africa: Summary Proposal for a Regional Technical Support Project – Phase 1". The project would assist participating countries to prepare a country framework document and formulate national statistical capacity building projects. It would later support implementation. We are, at present, seeking donor funding in support of this work.

Acknowledgements

Significant inputs from Robert Townsend, Karen Brooks, and Ernst Lutz are gratefully acknowledged

References

Aplers, E. (1977), "The East African Slave Trade", in Konczacki Z. and J. Konczacki (eds.), *An Economic History of Tropical Africa: Volume One—The Pre-colonial Period,* Totowa, N.J.: Frank Cass.

Binswanger, H. and E. Lutz (2000), "Agriculture Trade Barriers, Trade Negotiations, and the Interests of Developing Countries", paper presented at the High-Level Round Table for the United Nations Conference on Trade and Development, February, Bangkok.

Binswanger, H., K. Deininger and G. Feder (1995), "Power, Distortions, Revolt, and Reform in Agricultural Land Relations", in Srinivasan T. and J. Behrman (eds.), *Handbook of Development Economics*, Vol. III., 2659-2772.

Bloom, D. and J. Sachs (1998), "Geography, Demography, and Economic Growth in Africa", *Brookings Papers on Economic Activity*, 2.

Brunetti, A., G. Kisunko and B. Weder (1997), "Institutional Obstacles to Doing Business: Region-by-Region Results from a Worldwide Survey of the Private Sector", Policy Research Working Paper 1759, World Bank, Washington, D.C.

Coetzer, J., Thomson G. and R. Tustin (eds.) (1994), *Infectious Diseases of Livestock with Special Reference to Southern Africa,* Cape Town, South Africa: Oxford University Press.

Collier, P. and H. Binswanger (1999), "State Reconstruction, Civil Wars and Ethnic Conflicts", paper presented at the Can Africa Claim the 21st Century? Seminar Series, 6–10 July, Abidjan, Côte d'Ivoire.

Collier, P. and J. Gunning (1997), "Explaining African Economic Performance", Working Paper 21, Oxford University, Centre for the Study of African Economies, U.K.

Curto, J. (1992), "Historical Demography and the Effects of the Slave Trade in Africa: An Analysis of the Major Quantitative Studies", CDAS Discussion Paper 77, McGill University, Montreal, Canada.

Delgado, C., Hazell P., Hopkins J. and V. Kelly (1994), "Promoting intersectoral growth linkages in rural Africa through agricultural technology and policy reform", *American Journal of Agricultural Economics 76: 1166-1171.*

Donovan, G. and F. Casey (1998), "Soil Fertility Management in Sub-Saharan Africa", World Bank Technical Paper 408, Washington, D.C.

Fage, J. (1977), "Slavery and the Slave Trade in the Context of West African History", in Konczacki Z. and J. Konczacki (eds.), *An Economic History of Tropical Africa: Volume One—The Pre-colonial Period*. Totowa, N.J.: Frank Cass.

Goreux, L. (2001), "Conflict Diamonds", Africa Region Working Paper Series, Number 13, World Bank, Washington, D.C.

Hayami, Y. and J. Platteau (1997), "Resource Endowments and Agricultural Development: Africa vs. Asia", *Serie Recherche (Belgium)* 192: 1–54. Facultes Universitaires Notre-Dame de la Paix, Namur, and Faculte des Sciences Economiques et Sociales, Cashiers.

Johnston, B. and J. Mellor (1961), "The Role of Agriculture in Economic Development", *American Economic Review, 51 (4): 566-93.*

Josling, T. (1998), "Agricultural trade policy: completing the reform", Institute for International Economics, Washington, D.C.

Karshenas, M. (1998), "Capital Accumulation and Agricultural Surplus in Sub-Saharan Africa and Asia", African Development in a Comparative Perspective Study 1, UNCTAD, Geneva.

Lipton, M. (1993), "Urban Bias: Of Consequences, Classes and Causality", *Journal of Development Studies,* 29: 229–58.

Maddison, A. (1996), *Monitoring the World Economy: 1820-1992,* OECD, Paris.

McCarthy, F., Wolf, H. and Y. Wu (2000), "Malaria and Growth", Policy Research Working Paper, World Bank, Washington, D.C.

McLean, K., G. Kerr and M. Williams (1998), "Decentralization and Rural Development: Characterizing Efforts of 19 Countries", World Bank, Washington, D.C.

Meillassoux, C. (1981), *Maidens, Meals and Money: Capitalism and the Domestic Community,* Cambridge University Press, New York.

Mellor, J. (1966), *The economics of agricultural development*, Cornell University Press, Ithaka, N.Y.

Mellor, J. (1976), *The new economics of growth: A strategy for India and the developing world*, Cornell University Press, Ithaka, N.Y.

North, D. (1989), "Institutions and Economic Growth: A Historical Introduction", *World Development* 17 (9): 1319–32.

Rueschemeyer, D., Stephens, E. and J. Stephens (1992), *Capitalist Development and Democracy,* Chicago, Ill.: University of Chicago Press.

Sachs, J. and A. Warner (1998), "Sources of Slow Growth in African Economies", *Journal of African Economies*, 6(3):335-76.

Schiff, M. and A. Valdés (1998), "Agriculture and the macroeconomy", Policy Research Working Paper, Development Research Group, The World Bank, Washington, D.C.

Schultz, T.W. (1964), *Transforming Traditional Agriculture,* New Haven, Connecticut, Yale University Press.

Thirtle, C., van Zyl J. and N. Vink (eds.) (2000), *South African agriculture at the crossroads: an empirical analysis of efficiency, technology, and productivity,* St. Martin's Press, New York.

Tilly, C. (1990), *Coercion, Capital, and European States, A.D. 990–1990,* Cambridge: Basil Blackwell.

UNAIDS (1998), *AIDS Epidemic Update: December 1998,* Geneva.

UNCTAD (1998), *Trade and Development Report 1998,* Geneva.

Voortman, R., Sonneveld, B. and M. Keyzer (1998), "African Land Ecology: Opportunities and Constraints for Development", African Economic Research Consortium, Nairobi, Kenya.

Farm Households, Natural Resources and the Environment

Robert E. Evenson
Director, Economic Growth Center, Yale University.
e-mail: evenson@iiasa.ac.at

Abstract: Economic development policy requires better understanding of rural households and their responsiveness to economic changes. The agricultural household model offers guidance to more relevant data collection. Better Time Use data, in particular, are required. Developments in land, labor and financial contracts and market modeling also suggest directions for data. The increased importance of environmental and natural resource issues in agriculture also calls for better data documenting both investments in land and water resources and depreciation / degradation of these resources. It is important that data allow a better distinction between the producer good and consumer good dimensions of natural resources.

Keywords: Agricultural households; Land contracts; Labor contracts; Time use surveys; Full income; Land investments; Land degradation

1. Introduction

The economic development literature is placing new emphasis on economic growth and its measurement. This is, in part, because, contrary to some studies claiming that "trickle-down" effects are minimal, evidence suggest that most poverty reduction is achieved through growth. The "Kuznets Curve" effects, which suggest a worsening of income distribution as income levels rise from very low levels to be followed by an improvement in distributional parameters as income reaches high levels, have not been realized. Virtually all increases in real per capita income are accompanied by constant or improving income distribution.

The *operational issue of economic growth achievement* is thus back on center stage. The new endogenous growth literature offers models of growth that provide new insights regarding some aspects of growth but remain weak on the mechanics of growth convergence, i.e. of faster growth by economies with lowest incomes. For agriculture, there is now a general recognition that the high degree of location specificity of most technology used in the sector is the key barrier to technological convergence, where one country benefits from gains made in other countries. But, the question of transaction costs in rural markets has also become more important and serves as a focus for studies of convergence failure.

A second set of issues affecting the agricultural sector is *the heightened concern with the environment and natural resources*. Perhaps the central feature of this concern from a growth perspective is the recognition that environmental goods can be “producer goods” that produce growth or “consumer goods” that do not produce growth, or they can be both producer and consumer goods. The failure to distinguish between producer good and consumer good aspects of natural resources can result in serious impediments to growth. For example, if international organizations pressure developing countries to conserve and protect environmental producer goods, this may be consistent with growth achievement and poverty reduction. But, if they insist on imposing the “consumer tastes” of the rich countries on poor countries regarding environmental consumer goods, they can inhibit growth, in effect condemning poor countries to remain in conditions of mass poverty.

Environmental agencies are predisposed to argue that all environmental goods are producer goods and that they really do affect productivity and growth. For example, we see many arguments that farmers in poor countries should not have modern crop varieties because this will result in a loss of “biodiversity”. But if biodiversity is primarily a consumer good valued most highly by the rich, this argument has serious welfare implications for the poor.

Most *statistical systems* producing estimates of crop and animal production, land use, irrigation, input use, trade and prices were established many years ago. Changes in these systems have been made very slowly, even when changes have been called for by obvious discrepancies. The incorporation of changes in statistical systems to accommodate new concepts and increased political importance of activities associated with agriculture has been negligible. In this paper, I will discuss the possibilities for generating data to accommodate the conceptualization of the agricultural household model and to accommodate the increased emphasis and importance placed on natural resources associated with agricultural production.

2. Agricultural households and full income measurement

The agricultural household model is now widely accepted as an analytic structure for addressing a number of economic issues. The central feature of the model is that farm households produce non-marketed goods that are of significant value. But perhaps of more relevance is the fact that the same goods may be considered to be marketed goods in some households and circumstances and non-marketed goods in other households and circumstances in conventional measures of income.

The classic case of home-produced goods that are inconsistently treated in typical statistical systems is the service associated with the preparation of meals. When meals are prepared in households, these services are not treated as produced goods and do not generate income. When the same services are produced in a restaurant, they are treated as produced goods. Services provided for household upkeep and for childcare are examples of other home-produced goods that are inconsistently treated in standard data systems. (Actually, some systems do measure home-produced food preservation of canned goods, etc., but this is not done systematically).

Home-produced goods also include the production of farm capital goods, for example in the form of maintenance of drainage and irrigation systems, and these have implications for natural resource measurement (see the next section).

2.1 Policy relevance of full income measurement

Why would it be important to have better measures of home production?

Perhaps the most important reason is to obtain more consistent measures of income growth. Full income can be expressed as the sum of the value of home-produced non-marketed goods plus the value of marketed goods (including home-consumed marketed goods) produced by the household (net of purchased factors of production and imputed capital services). Surveys for low income countries (e.g. the Philippines) indicate that full income is approximately double the income as conventionally measured (i.e. home-produced non-marketed goods are of approximately equal value to the net income as conventionally measured). For high-income countries (e.g. the USA), full income is less than 50 percent higher than conventionally measured income (Florencio *et al*., 1979).

This difference is due to several factors. First, many home-produced goods are associated with children, and family sizes are lower in high income (high wage) economies. Also, as incomes rise, families spend more on marketed goods that substitute for home-produced goods (e.g. restaurant meals). In addition, as wages rise and market costs fall, the comparative advantage in producing goods in the home changes in favor of market-produced goods.

Clearly, for purposes of measuring real income growth, growth in full income is the conceptually correct measure. Given the pattern noted above, growth in full income will be slower than growth in conventionally measured income. In a rapidly growing economy (e.g. Korea), conventional income may grow at 10 percent per year, while real full income is actually growing at eight percent per year (Evenson and Roumasset, 1986).

A second reason for measuring full income is that the difference in full income and conventional income is informative regarding the shift of activities from household enterprises to market enterprises. Understanding this process is important for policy making.

A third reason for measuring full income is that it provides a more accurate gender perspective. Most home-produced income is produced by women. Measuring these activities more carefully is important for understanding gender issues (Mwabu and Evenson, 1996).

2.2 Measurement techniques

As with conventional income, the home production component of full income can, in principle, be measured in two ways: (1) as the sum of the value of home goods produced, or (2) as the sum of the value of the resources used to produce those goods.

Since many of the home goods are not marketed and since their "shadow prices" are actually endogenous to, and thus may differ by households, it is not feasible to measure the value of home goods directly (first approach). It is, however, feasible to value these goods by valuing the resources used to produce them (second approach) because the chief resource used is time. Time use surveys can thus be used to measure home production (Evenson, 1983).

There is considerable experience with time use surveys and the methods are such that they can be incorporated into statistical data systems. Typically, this would entail implementing time use surveys for sub-samples of households. These can include agricultural households and other types of households.

One of the activities in developing countries that is poorly understood is the household enterprise. Often rural households engage in simple manufacturing and fabricating activities as well as the more traditional shop keeping and home production enterprises. Many household enterprises, however, provide the experience and entrepreneurial training that leads to more formal industrial development (e.g. the Village Enterprise System in China).

Another policy theme that is emerging in developing countries is the issue of structural change. The structure of agricultural production encompasses farm size, production specialization, off-farm employment and household enterprises. We know that, in developed countries, a major structural change is taking place. Farm size is growing, production specialization is increasing: in fact, poultry and egg production is effectively an industrial enterprise. Pork production is also being industrialized and dairy production is, as well. The number of farmers and farm workers is rapidly declining as agriculture has become more capital intensive. Studies show that relative prices affect these changes and that technology does, as well.

As these changes become more important in developing countries, it will be important to understand how farm households deal with them. History has shown that rapid structural changes can impose significant costs on agricultural households.

Policy makers in developed countries have found it difficult to deal with structural change, particularly where large labor force and occupational changes are required. While policy makers know that large changes are almost impossible to avoid, many governments have been willing to spend large sums to avoid rather than facilitate the adjustments required. It is difficult to serve an interest group through policies that reduce the size of the interest group.

This is exacerbated by land valuation issues. The “farm crisis” in the U.S. of the 1980s was greatly worsened by irrational valuation of farm property leading to excess credit. As more governments face these adjustments, it is important that statistical time series monitor changes, especially in the labor force, that are taking place.

Another topic in the household economics area is an old topic, the measurement of food consumption. Reviews of food consumption data have consistently noted that better measures are called for (Evenson and Pray, 1994). In a recent book (Smil, 2000), Vaclav Smil is particularly critical of food consumption measures and argues that more attention should be given to measuring food wastes. He also finds the caloric consumption data collected by FAO and other agencies to be misleading. Other critics of food consumption data have also made these points. Lipton (FAO, 2000), argues that a strong correlation does exist between food production and consumption per capita. This conclusion reinforces the need for a review of food consumption data.

3. Natural resource capital measurement

Agricultural land and water resources and associated climate elements constitute an agricultural capital stock of considerable value. This stock can be increased by investments and good management practices. It can also be depleted by poor management practices and natural events (e.g. some soil erosion is unavoidable even under best management practices).

Actually, many countries have soil and water conservation programs designed to improve management practices, and many farmers invest in drainage systems and irrigation systems to build up this capital stock. Yet, most current interest in soil and water capital ignores these investments. A recent study of agricultural productivity in U.S. agriculture (Acquaye *et al*., 2000) showed that land area in U.S. agriculture decreased from 1.43 billion acres in 1949 to 1.17 billion acres in 1991, an 18 percent decline. However, the real productive capital stock value per hectare rose by more than six percent so the real land stock decreased by 12 percent. Some of this land improvement was through investment in drainage, some through investment in irrigation and some through improvements in management. The Conservation Resource Program acreage was subtracted from the stock. Cash rental values were used to make quality adjustments.

It is important that data on productivity change be linked to measures of natural resource capital change to enable studies of real producer goods effects. For example, studies of soil erosion may show what observers would regard to be high rates of erosion. But the important policy issues are associated with the erosion-productivity link. If erosion reduces productivity, programs of erosion control may be justified. If productivity effects are minor, spending large sums to prevent erosion will result in reduced welfare for the people in the economy (even when it is perceived to be worth doing by international environmentalists' groups).

If statistical systems are to honestly value the natural resource capital in agriculture, it is important that they treat valuation seriously and measure investments and land-water improvements as well as losses through soil degradation and related damages to this capital stock.

Many policy makers concerned with natural capital reveal a strong predisposition to consider family farmers to be poor custodians of natural capital. Yet, household studies, including studies of "slash and burn" households, show that farmers are stewards of natural resource capital. The actual degree of stewardship does depend on property rights and on public programs. Actual changes in the value of natural resource capital differ according to climate conditions, the extent of population pressure on land, on prices and on crop technology. The yield increases due to genetic improvement, and the consequent low cereal grain prices, reduced incentives in most countries to expand cultivated land, even in the presence of historically unprecedented population increases.

Better statistical data on land use, land quality, land rents (e.g. cash rents) and drainage investments will be valuable for policy makers. This is in addition to measures of soil degradation by region and of related natural resource capital problems. Conservation practices can also be monitored.

A number of countries incorporate relevant questions in their Agricultural Census systems. For example, Brazil, from 1970 to 1995, collected data on land values and on investments in irrigation. These data have been valuable for studies of climate change but could be used to monitor natural resource capital stocks (Sanghi *et al.,* 1998).

Time use and household enterprise surveys can also include activities that are relevant to natural resource capital. These include enterprise production of capital, drainage, irrigation ditches and maintenance activities. Ideally, we should be able to match survey data with existing data on soil types and classes. This can be accomplished by sampling design techniques.

One of the major policy issues of concern to agriculturalists today is climate change. Climate is effectively part of the natural capital stock in agriculture. A number of studies have attempted to utilize this to infer what would happen if temperatures were to rise and rainfall levels and variabilities were to change. These studies have use “Ricardian” evaluation techniques to associate the value of natural resource capital with climate components. This requires data on land values, specifically on the value of land distinct from the value of buildings and other reproducible capital. Only a few agricultural censuses allow this kind of valuation, although there are efforts of use proxies such as “net revenue per hectare” for these purposes.

The climate change studies to date are indicating disturbing prospects for climate change. They suggest that rising temperatures will benefit many regions where temperatures are relatively low and will harm those regions with high temperatures. Since the regions most likely to be harmed are the tropical poor countries, climate warming appears likely to exacerbate the already widening difference in incomes between rich and poor countries.

Valuation of natural resource capital in agriculture, then, is important for a number of policy issues. And these policy issues have staying power in that they will become more important in future years.

Biodiversity in agriculture is of growing concern to many policy makers. For a number of reasons, including for the measurement of biodiversity, surveys of varietal use can be valuable. These surveys would enable measures of varietal turnover that could then be related to research programs. They would enable better monitoring of dynamism in agriculture. And they would allow better measures of biodiversity change.

As with soil erosion and other environmental goods, the real producer good aspect of biodiversity is the effect on production and productivity change of a change in biodiversity. It is vital that the productivity-biodiversity link be established with data because advocates of environmental consumer goods generally do not consider the damage to local welfare of imposing the tastes of the rich on poor people.

4. Summary

Economic growth achievement and associated poverty reduction has varied greatly between developing countries. A number of countries have real per capital income that is similar or below levels of four or five decades ago. Other developing countries have achieved historically unprecedented rates of economic growth and poverty reduction over these decades. World food production has increased more rapidly than population globally and real international prices for basic food grains are at all-time lows. Yet, large numbers of people remain inadequately fed.

Agricultural statistics systems are also grounded in the early post-war period when they were established in many countries. These systems have shown little responsiveness to changes in the policy environment. They also have been unresponsive to criticisms regarding data discrepancies.

In this paper two sets of policy concerns, the first dealing with agricultural households, the second with natural resources, are discussed and the implications for modifications and extensions of statistics systems are discussed. However, these extensions are probably of less importance than the need for statistics systems to address their most glaring problems of inconsistency and inadequacy in measuring production, land area and inputs.

The agricultural household model is now widely accepted in the economic development literature as a useful tool for understanding the behavior of farm families. A related body of literature on land, labor and financial contracts is also relevant for extensions to data collection systems. The most glaring need is for more consistent measurement of household income. The "full incomes" concept of the agricultural household literature includes income from home-produced non-marketed goods. Time use surveys can provide measures of this income. Additional data on land contracts, water contracts, labor contracts and agricultural finance can also be obtained, especially in Censuses. Researchers are beginning to note that households also engage in a number of important enterprises, many of which provide valuable experience for industrial development.

There is little question that the concern with environmental goods is of increasing importance. This concern is often marked by a failure to distinguish between producer goods that are vital to economic growth and consumer goods that are not (in fact environmental consumer goods are dependent on income, i.e. high income elasticities). Environmental groups are pre-disposed toward treating resource-based goods as producer goods (to avoid the charge that they are imposing the tastes of the developed countries on the developing countries). This leads to serious bias in measurement (e.g. measuring soil degradation but not soil improvement).

Statistical systems have the potential to reduce the range of disagreement between interest groups and to foster objectivity in policy debates. *Numbers do have power*. It is important that official statistics systems respond to both their critics regarding existing data and to shifts in policy concerns.

References

Acquaye, A., J. Alston, and Ph. Pardey (2000), "A Disaggregated Perspective on Post-War Productivity Growth in U.S. Agriculture: Isn't that Spatial?", presented at the NC-08 Conference on *Agricultural Productivity: Data, Methods and Measures*, USDA-ERS, Washington D.C., March 9-10, 2000.

Evenson, R. (1983), "The Allocation of Women's Time: An International Comparison", *Behavioral Science Research,* 17 (304), Spring-Summer 1983.

Evenson, R. and C. Pray (1994), "Measuring Food Production (with Reference to South Asia)", *Journal of Development Economics*, 44, Elsevier Science.

Evenson, R. and J. Roumasset (1986), "Markets, Institutions and Family Size in Rural Philippine Households", *Journal of Philippine Development*, Vol. XIII.

FAO (2000), *The State of Food and Agriculture, 2000*, Rome.

Florencio, C., B. White, H. Binswanger and R. Evenson (eds.) (1979), *Rural Household Studies in Asia*, University of Singapore Press.

Mwabu, G. and R. Evenson (1996), “Household Composition and Expenditures on Human Capital Formation in Kenya”, *Research in Population Economics*, Vol. 8, 205-232, JAI Press.

Sanghi, A., A. Dinar, R. Mendelsohn, J. Parikh, K. Kumar, J. McKinsey, S. Lonergan and R. Evenson (1998), “Measuring the Impact of Climate Change on Indian Agriculture”, World Bank Technical Paper No. 402, The World Bank, Washington D.C.

Smil, V. (2000), *Feeding the World*, The MIT Press, Cambridge, Massachusetts.

The Response of Agricultural Statistical Systems to the New Demands

Pratap Narain
Senior Officer, Statistics Division, FAO, Via delle Terme di Caracalla, 00153 Rome, Italy, Tel: +39 06 57055577
email: pratap.narain@fao.org

Abstract: Two major new issues, namely, the need for a comprehensive approach for environmentally sound and sustainable economic development and the right to have access to safe and nutritious food, were raised at three important World Summits organised in the last decade. As these issues are directly as well as indirectly related to agriculture, the system of agricultural statistics would need to provide appropriate information for decision making. In this connection, it is necessary to integrate data on natural resources and environmental issues with the other databases by adopting suitable survey techniques. Furthermore, a major consideration would have to be given to re-organising the architecture of the international and national statistical systems.

Keywords: Sustainable economic development; Food security; Data integration.

1. Developments over the last decade

It is a well-known fact that two-thirds of the population in developing countries live and work in rural areas where the principal economic activity is agriculture. For quite some time the importance of agriculture has been viewed primarily from the perspective of the world's growing population and the consequential pressure on land resources and, hence, on society's capability to adapt to the increasing needs relating to food, fibre and shelter. These needs led to an agricultural statistics system that focussed on issues relating to production and productivity. Statistical techniques were refined and improved estimates of agricultural production were generated to understand the "what, when and where" of the problem.

In the last decade, however, two major new issues that are directly and indirectly related to agriculture surfaced. The first issue which was debated in the seventies and eighties, was raised at the 1992 United Nations Conference on Environment and Development (Earth Summit) in the form of Agenda 21. This agenda highlighted the need for a comprehensive approach for environmentally sound and sustainable economic development. The second issue was recognised at the 1995 Copenhagen World Summit for Social Development (Social Summit) and the 1996 Rome World Food Summit (Food Summit). At these Summits it was recognised that everyone has a right to have access to safe and nutritious food. The Food Summit declared that "Poverty is a major cause of food insecurity and sustainable progress in poverty eradication is critical to improve access to food. Conflict, terrorism, corruption and environmental degradation also contribute significantly to food insecurity. Increased food production, including staple food, must be undertaken. This should happen within the framework of sustainable management of natural resources, elimination of unsustainable patterns of consumption and production, particularly in industrialised countries, and early stabilisation of the world population." Countries have the primary responsibility for creating an economic and political environment that assures the sustainable maintenance of the global ecosystem and provides food security to their citizens. Agriculture is at the centre of both of these issues. Agricultural statistics systems need to be geared up to meet these new demands. This paper discusses data that are needed to formulate policies and programmes in the light of recommendations made by these Summits and what steps need to be taken to improve the existing statistical systems. While the paper takes a comprehensive view of agriculture (including forestry and fishery) in general, only crop and animal husbandry have been considered in the section on the new approach to the data collection.

2. What are the future data needs ?

Future policy decisions relating to agriculture should be taken at the meso level by considering people, institutions and the ecosystem in smaller homogeneous groups. Any action and policy formulation would require to take into account not only the internal situation but also their impact on the neighbours and the world globally. To understand the future responsibility of an agricultural statistics system more clearly, it is necessary to go in further details of the recommendations that have been made by the Summits.

2.1 The 1992 UN Conference on Environment and Development (Earth Summit)

Agriculture has an important place among various economic activities, not only because it is a principal user of natural resources but also because it is a user as well as a producer of eco-products (marketable products, recreation, maintenance of species, environmental quality, etc.). Per capita arable land has shown a progressive decline associated with increasing population density in rural areas. Like land resources, water resources are not evenly distributed and several countries face scarcity constraints. The intense pressure on land and water resources often leads to their misuse. The area under cultivation is being expanded by cultivating marginal lands including sloping lands, and lands which have been deforested. To obtain higher production levels, measures are being taken to improve yields per hectare but these involve the application of heavy dosages of agricultural chemicals, and expansion of irrigation, which, unless managed properly, can have adverse environmental consequences. Thus, the extensive use of fertilisers is often linked to eutrophication of water bodies, soil acidification and contamination of the water supplies with nitrates. The problem of excessive use of nutrients is confined to developed countries and some developing countries only. In fact, in many developing countries, nutrients being applied per hectare are insufficient to replace those that are removed by crops. The increase in the use of pesticides in response to rising pest problems is also having a negative impact on the environment. The management of water resources too is posing problems in many areas. In water scarce areas, water withdrawn from aquifers is accompanied by a relatively low rate of recharge. This puts a tremendous stress on available water resources. In several other areas where irrigation is practised without adequate drainage, over-watering often results in waterlogging and salinization. The impact of climate change on agriculture and the ecosystem is becoming an additional source of concern.

The Earth Summit was convened to address such urgent problems of environmental protection. Agenda 21 addresses the pressing problems of today and also aims at preparing the world for the challenges of the twenty-first century. It reflects a global consensus and political commitment on environmental development. National strategies, plans, policies and processes are crucial in achieving this. The summit recognised that, although countries have the sovereign right to exploit their own resources pursuant to their own environmental and developmental policies, they cannot cause damage to the environment of other States or of areas beyond the limits of national jurisdiction. They have therefore to co-operate in a spirit of global partnership to conserve, protect and restore the health and integrity of the Earth's ecosystem. The implementation of the developmental and environmental objectives of Agenda 21 will require a substantial flow of new and additional information / data on natural resources.

Environmental statistics are of a multidisciplinary nature. They include a large set of physical databases on climate (solar energy and light, water, wind, heat and temperature), flora, fauna, etc., which are collected using various techniques and methods for a comparative analysis of their impact on the ecosystem and the social, demographic and economic conditions of the people. In order to achieve sustainable development, environmental protection shall constitute an integral part of the development process and cannot be considered in isolation from it.

2.2 The 1995 Copenhagen World Summit for Social Development (Social Summit) and the 1996 Rome World Food Summit (Food Summit)

At the World Food and Social Summits it was recognised that world population is growing and steps need to be taken urgently for eradicating hunger and malnutrition. The World Food Summit calls for determined economic policies that will permit producers of agricultural products to adopt appropriate input technologies, farming techniques, and other sustainable methods. The Summit also stressed that production increases need to be achieved without further overburdening natural resources. Establishing sustainable and diverse patterns of production should take into account the present and future needs of the people as well as the natural resources potential and limitations. It was recognised that developing countries are generally poorly equipped in terms of technical and financial resources and educational infrastructure, in particular in rural areas. In these areas, lack of income opportunities, failure of crops due to flood and drought, inadequate commodity (inputs and consumer goods) distribution networks, limited access to public services and the poor quality of these services are all fundamental aspects that need to be considered with regard to rural food security. Generally such socio-economic conditions and ecosystem degradation create a vicious circle. Therefore, it was recognised that the policies that provide an effective incentive structure for sustainable management of natural resources will help ensure that national agricultural practices are developed and implemented in a holistic approach.

The framework for the development of environmental statistics (FDES) recommended by the UN (UN, 1991) envisaged this need and related the components of the environment categories with the other information categories. The components of the environment categories consist of the data on flora, fauna, atmosphere, water, land/soil that have been combined with the data on the man-made environments (human settlements). The information categories, which are based on the recognition that environmental problems are the result of human activities and natural events, are (i) social and economic activities, and natural events; (ii) environmental impacts of activities and events; (iii) responses to environmental impacts and events; (iv) inventories, stocks and background conditions. Thus, having recognised the multifaceted character of the issues raised by these Summits, the following data (in addition to the usual data on production, productivity, etc.) would be required to formulate the appropriate policies and monitor the progress made:

- Data on socio-economic conditions of agricultural / rural households (access to means of production such as land, water, inputs, improved seeds and plants, farm credit, etc.).
- Data on rural infrastructure to effect timely transfers of supplies to deficit areas.
- Data on climate, incidence of drought and desertification, pests, erosion of biological diversity, and degradation of land and aquatic-based natural resources including water and watersheds in depleted and overexploited areas to combat environmental threats to agriculture and achieve greater production as well as to restore and rehabilitate the natural resource base.
- Data on trade, marketable surplus, prices and cost structures to promote appropriate trade policies to enhance sustainable agricultural growth.
- Data on mixed-farming systems, organic farming, etc., to assist such farming operations to become profitable.
- Data on the processing and marketing of diverse food products and by-products, in response to the needs of the consumers for properly balanced diets.

The data listed above can be broadly classified into three groups, viz., socio-economic data collected using conventional methods, socio-economic data collected using a mixture of conventional methods and remote sensing and technological and scientific data. Each of these groups has a different role as well as limitations in developing a statistical system.

3. The current state of the food and agricultural statistics systems

The statistical system of any country has usually been developed over the years primarily to meet the needs of the country's administrative set-up. In any statistical system including the agricultural statistical system, part of the statistical data originate from information contained in various official administrative returns submitted to the government, while another part of data are collected through sample surveys and censuses organised by the statistical services of the administration to meet specific needs. In view of the future needs as discussed above, agricultural policy formulation and decision making would also need data on the state of the natural resource base and environment. With technological breakthroughs in the field of information science, satellite based technologies like remote sensing techniques form another important source of data which may be different from other socio-economic data sets. From the point of view of uses of these data sets, we can group them into two sets, namely: (i) non-georeferenced databases, consisting of statistical information on human activities including agricultural production, land use, etc., and (ii) geo-referenced databases, consisting of physical databases on the distribution and availability of natural resources such as soils, climate, water, flora,

fauna, etc., and databases of environmental monitoring activities originating from monitoring sites.
Future decision making processes would need to integrate both types of data in modelling format for making plans and policies. In fact, a number of decision support tools are already being developed to perform this function in a rather restrictive manner. In the future a more comprehensive view would be required. The current status of these two types of data are presented in the section below.

3.1 Non-georeferenced databases

The availability of food and agricultural statistics in developing countries varies from country to country and is highly correlated with a country's level of development. Even within any given country, the various components of the agricultural statistics system are typically at different levels of development. In Kabat *et al*. (1998), it was recognised that ".... Despite the increasing awareness of the importance of statistics in planning for agricultural development, it is noted that most developing countries still do not have an adequate system of statistics pertaining to the agricultural sector. The available agricultural data are incomplete in terms of (a) the range of commodities covered (for example, in many cases only cash crops for large farms are covered), (b) the range of variables or data sets covered (for example data on agricultural inputs are in many countries practically non-existent), and (c) coverage of the nation (sometimes parts of the country are excluded from the national statistical reporting system). Furthermore, even when data are available, their reliability is often questionable. In fact, agricultural data in a country also come from other kinds of censuses or surveys as well as from administrative records. As the different institutions involved are not aware of each other's activities, there is considerable duplication of effort and, in many cases, conflicting data are reported for the same items. As regards data on the welfare of the households engaged in agricultural activities and the food consumption level of the population, in many countries appropriate household surveys are either not available or not regularly undertaken. When it comes to data relating to the depletion of land resources and their environmental effects the situation is worse. Finally even when considerable data relating to the agricultural sector are available, it has seldom been recognised that the different data components often have different coverage and time frames thus requiring special processing, tabulations, adjustments etc. prior to their usage in an integrated manner or for the purpose of a particular study or analysis. The reasons for this state of affairs are many." The paper suggests a Comprehensive Approach to deal with these issues.
This review of 'issues and concerns' for the statistical data in developing countries was made from the limited perspective of the data producer. When one talks about future needs from the data user or from the analytical perspective, it would be useful to understand that in the coming time it would be necessary to integrate not only the various aspects of the economy but also the data to review the process of development in comparison with the rest of the world or neighbouring countries. This requirement calls for two essential features to be embodied in the data sets, viz. the adoption of uniform concepts, definitions and classification systems and the creation of a continuous time series. The latter is also important, as disjointed data series from different domains or countries can not be aggregated.

3.2 Geo-referenced databases

In recent years, remote sensing has become an important major source of geographic information relating to natural resources and the ecosystem. The information that is available at global, regional and local level is being used for early warning and environmental monitoring, rapid assessment of crop growing conditions, land cover, forest resources assessment, generation of agro-climatic databases, etc. As these data sets are project specific, it is virtually impossible to assemble a comprehensive list of the databases. An illustrative list of such databases can be seen at the FAO website. Recognising that most sustainable development decisions are cross-sectoral in nature and most of the natural resource information development agencies are single sector oriented, Geographic Information System (GIS) technology can help establish cross-sectoral communication by providing not only very powerful tools for storage and analysis of multisectoral spatial and statistical data, but also integrating databases of different sectors in the same format and structure for easy comprehension.

4. Suggestions for a comprehensive agricultural statistical system that meets future needs

A statistical system is a link between data owners and data users. Data owners in the present context are the institutions. The system consists of managers, i.e. statisticians who collect, compile and disseminate the statistical data and other information as per the direction laid down in institutional arrangements. The role of statisticians has been critically examined by an Ad-Hoc Committee appointed by the ISI (ISI, 1992).

Many of the problems and constraints associated with food and agricultural statistics are typical of the development of national statistical systems in general and have been identified and discussed at various fora such as the Addis Ababa Plan of Action for Statistical Development in Africa in the 1990s (UN/ECA) and the 1999 FAO/World Bank/USDA Workshop on strengthening national systems of food and agricultural statistics in Africa. A number of factors have been identified such as poor co-ordination of scarce resources (especially of donor assistance), programmes reflecting the interests of donors rather than those of the country, non-sustainability of externally funded programmes, lack of trained manpower, high turnover of staff from statistical systems and poor management practices, technical and methodological constraints, poor data integration from various sources and a lack of 'value added'. At these fora, suggestions have been made for improving institutional arrangements and co-ordination among the data producers, firm commitments from governments, etc. However, these suggestions once again were directed to improve the existing agricultural statistical system. To meet the future needs of integrated data analysis it would be essential to look at basic architecture of the statistical set-ups at the national and international level with the objective of (i) to improve official statistics and its international comparability, (ii) to achieve closer co-ordination among the national and international organisations and (iii) integration of environmental and other economic databases. Systematic details about designing statistical information systems for national and international statistical organisations covering survey processing system, data warehouses, data reporting system (raw data, metadata, electronic document) and other implementation aspects, have been presented in UNECE (1999). In the present context, actions would be required by the national governments as well as the international community that go beyond statisticians and statistical offices. It would be necessary to incorporate changes in the data collection, processing, storage as well as dissemination systems. Some of these issues are highlighted below.

4.1 Actions required at the national level

Apart from what has already been said at different fora, in the present context, it would be important to focus on the issues relating to data integration for creation of an information system. Looking at the nature of the data originating from the three types of sources and the needs of decision makers it is obvious that it is essential to integrate these data sets. Integration of socio-economic data collected in statistical inquiries with the other derived secondary statistics and the information collected from administrative sources are being practised by most of the statistical offices. What is new is the integration of physical and monetary data with scientific and technical data produced by remote sensing and other systems. The utility of the national accounting framework in integrating various data sets is well known. However, there are certain limitations in what can be accommodated in national accounting framework (even when it is extended to cover environmental issues) as the accounting structure is based on the monetary system. However, in several cases additional or different requirements necessitate the development of complementary or alternative categories and concepts for collection of data. The 1993 SNA (UN, 1993) realised the need to expand the analytical capacity of national accounting for selected areas of social concern in a flexible manner, without overburdening or disrupting the central system and recommended use of satellite accounts. The satellite accounts are useful in integrating physical and monetary data. On the one hand, satellite accounts are linked with the central framework of national accounts and through them to the main body of integrated economic statistics. On the other hand, as they are more specific to a given field or topic, they are also linked to the information system specific to this field or topic. In this manner it allows for better integration of monetary and physical data.

In a satellite account it is desirable that the concepts, accounting structure and classifications should be tailored to the specific analytical purposes which facilitate data integration. Satellite accounts can focus on the role of people or natural resources. This calls for an extra breakdown of the economic sector for conceptual and numerical linkages among all kinds of related monetary and non-monetary phenomena, which may be expressed in different measurement units. To understand the dynamic nature of the economic system in terms of the social, economic and demographic characteristics of the population, it is essential to integrate information and data on various aspects. This task requires some basic considerations for collection and compilation of statistical data. Kabat *et al.* (2000) discusses some guiding principles to achieve this in a cost-effective manner.

Another approach to integrate physical data with the data on natural resources is the use of the Geographical Information System (GIS). In this approach various data sets are integrated using sub-national spatial units. To integrate the basic socio-economic data collected in various statistical censuses and surveys with the other geo-referenced databases using the GIS, it would be desirable to use survey designs that are based on a combination of the area and list frames. In the recent past, FAO has issued guidelines for conducting census and surveys using multiple frame survey techniques.

Area sample surveys usually consider a subdivision of the territory into land-use strata. The strata are defined by intensity of cultivated land, predominance of certain crops,

special agricultural practices, average size of cultivated fields, agro-urban areas, or other land-use characteristics. The delimitation of strata is done using satellite images, mosaics of aerial photos, topographic charts and field controls. In the classic procedure for the construction of an area frame, stratum boundaries are physical terrain features (roads, paths, rivers, etc.), that could be located on the ground. The area sample designs consist of a one or two-stage stratified replicated and self-weighted probability sample of "segments" (small land area). Socio-economic databases using a combination of area and list frame can easily be integrated with other geo-referenced databases.
An additional point, which needs to be considered, is to use multi-stage survey designs in a special way. For example, if at the first stage of selection administrative blocks or watershed are selected, information on some of the important items like infrastructure facilities, land use, etc., can be collected for understanding the depth of issues related to food security and causes of land degradation. At the second and third stages data relating to households and holdings can be collected for generating other information. This approach would provide data that could be integrated with other scientific and technical databases in a most cost-effective manner.

4.2 Actions required at the international level

At the international level, it would be necessary to prepare a long-term, integrated, user driven program for establishing an information system that caters to the national and international needs. To provide the harmonised information to different types of users data bank would be required at local (country), regional and global (international) levels. While countries would be required to generate and store national and sub-national level databases, regional and international organisations may organise similar data banks to store data as per their jurisdiction. Links may be established by incorporating meta-databases using a common classification such as presented in the Directory of International Statistics (UN, 1982). A few actions in this respect would be required:

- Strengthening existing work on standards and classifications to generate harmonized information that is compatible for comparison and aggregation purposes.
- With the requirement of extended data coverage, there would be substantial increase in the work load at country level. Steps may be taken at international level to develop unified questionnaires for collection of data to reduce respondent burden and harmonize concepts.
- To conceive an information system architecture which could be used to store a variety of data to draw information according to their need.

Some work is already going on in this direction. Reference could be made to the discussions in the last ACC Subcommittee on Statistical Activities (Washington D.C., September 2000 - Quality assurance, arrangements with common questionnaires and conflicting statistics) on generating unified questionnaires. Likewise, one can draw experience from the IMF in establishing "The General Data Dissemination System" (ESCWA, 1999) for collection of national accounts, government finances, balance of payment and other financial data or the Global Terrestrial Observing System (GTOS, 1998) jointly sponsored by FAO, International Council for Science (ICSU), UNESCO, UNEP and the WMO, for understanding the global change in the terrestrial ecosystem.

At the international level, a similar system may be established to collect and organise the data for future needs.

5. Concluding remarks

A forward looking agricultural statistics system that caters to the new demands would need to build on the motto: "look globally and act locally". For this purpose, the experts working in the fields of survey design, national accounts and geographical information systems would need to come together. The 'response' is hidden in the simple but difficult concept of co-operation. In short, it needs promotional efforts by the international organisations supported by financial support of donors.

Acknowledgements

The author would like to acknowledge the guidance and comments provided by Mr. L. Naiken, Chief, Statistical Analysis Service, FAO

References

ESCWA (1999), "The General Data Dissemination System by the International Monetary Fund", paper presented to the ESCWA Statistical Committee, Beirut.

FAO (1996a), "Report of the World Food Summit, Rome Declaration on World Food Security and World Food Summit Plan of Action", Rome.

FAO (1996b), "Multiple Frame Agriculture Surveys, Volume I", *Statistical Development Series* Number 7, Rome.

FAO (1998), "Multiple Frame Agriculture Surveys, Volume II", *Statistical Development Series* Number 10, Rome.

GTOS (1998), "Global Terrestrial Observing System Implementation Plan", December, 1998, FAO, Rome.

ISI (1992), "The Role of Statisticians", Report of the Ad-hoc Committee, *International Statistical Review*.

Kabat, L., L. Naiken, D. Marshall and P. Narain (1998), "Issues and Concerns for Developing Countries", paper presented at 'Agricultural Statistics, 2000 - an International Conference on Agricultural Statistics' jointly organised by ISI, USDA/NASS, FAO, Eurostat and OECD.

Kabat, L., L. Naiken and P. Narain (2000), "Data integration and its role in the development of better agriculture and food information systems", Invited paper presented at 'Integrating Agriculture and Food Statistics: National and International Perspectives', International Conference on the Establishment of Surveys – II, Buffalo, New York.

UN (1982), "Directory of International Statistics", *UN Statistical Papers*, Series M number 56 (Rev. 1), New York.

UN (1991), "Concepts and Methods of Environment Statistics: Statistics of Natural Environment", New York.

UN (1992), "Agenda 21, the Rio Declaration on Environment and Development", New York, (http://www.un.org/esa/sustdev/agenda21.htm).
UN (1993), "System of National Accounts", New York.
UN/ECE (1999), "Information systems architecture for national and international statistical offices: Guidelines and recommendations", *Statistical Standards and Studies* No. 51, UN Statistical Commission and Economic Commission for Europe, Geneva.

5 June 2001

Invited Paper Session

CHANGES IN SOCIAL ASPECTS OF AGRICULTURE

Chair: G. Calò

Family Farms: Are They in Danger of Extinction?

Carlo Russo
Università di Cassino, Via Mazzaroppi snc, Cassino (FR), 03043 Italia,
e-mail: carlorusso@eco.unicas.it

Massimo Sabbatini
Università di Cassino, Via Mazzaroppi snc, Cassino (FR), 03043 Italia,
e-mail: m.sabbatini@eco.unicas.it

Abstract: food system globalization is posing a serious threat to small operations run by family farmers. Statistical data suggest a potential crisis for the family farm model, at least in the more developed countries. Small farms, based on family labor are losing competitiveness compared to large, industrialized enterprises. However, a broader perspective is needed: in the near future, family farms will not face extinction if they are able to offer a set of social benefits in addition to food production. Food security, land preservation, environment protection, eco-turism are just a few examples. Agricultural census should be able to describe all the multiple contributions of family farms, in order to provide unbiased information to decision-makers. A brief analysis of 17 census questionnaires showed that, while production and land use data are collected extensively, the social role of family farms is often overlooked.

Keywords: Family farms, social role of agriculture

1. Introduction: family farms in the global food system

The global food system is undergoing rapid and radical changes. The WTO agreements have fostered a new era of free trade contributing to creating a global competitive environment, in which economy of scale, cost effectiveness and innovation are the most important competitive advantages. With world commodity prices falling, only the most efficient farms are likely to be able to compete successfully. Few data can explain the magnitude of the challenge: compared to 1996/97, in 1999/2000 export prices decreased by 28.1% for rice, 38.1% for wheat and 32.6% for maize (FAO 2001). In such competitive environment, farmers are exposed to potential reduction of income, because price decreases more than offset the average increases in yield. The pursuit of efficiency is one of the most relevant driving forces of the 21st century food market.

In this scenario, the so-called *industrialized agriculture* seems to be the fittest organizational model to compete on the world market. This model is based on large-scale, capital-intensive farming, which is able to deliver a large amount of agricultural commodities at a relatively low price. Family farms, unable to capture large scale economies and lacking the capital to implement new technologies, seem incapable of achieving the same yield levels and the same cost-effectiveness. As a consequence, family farms are progressively driven away from the world market. Recent studies have shown a significant increase in food export from developed countries (with industrialized farms) and a reduction of the developing countries' share, where traditional farms are still the backbone of agriculture (FAO 2001).

The decline of government support due to international negotiations and agreements within WTO is worsening the scenario. Without public contributions, family farms are even less likely to match the efficiency required by international competition. The import of low-cost food produced by industrialized farms is now affecting regional markets as well, and is challenging small farms on their own ground. In the past, social needs and food security concerns made governments worldwide sensitive to family farm issue and trade barriers were created. Now, international agreements limit public intervention and leave farmers more exposed to price fluctuations and income risk.

This scenario questions the role of family farms in the 21st century global food system. The competitive disadvantages are stressed: family farms would be less innovative, less efficient and unable to meet the food demand of the increasing world population. Moreover, traditional small farms are considered less able to meet the needs of global customers because of their capital constraint and their limited information. Thus, they might be left out of an evolving food system based on innovation, flexibility and consumer satisfaction. From this point of view, the marginalization of the role of the family farm may be considered as a serious threat.

In this paper, the competitiveness of family farms and their role in the evolving food system are discussed. The objective of the analysis is to offer a brief outline of the strengths and weaknesses of the family farm and a perspective of future developments. In order to achieve this objective, the economic and social role of family farms needs to be addressed, so that a complete evaluation may be made on their impact on local communities. The analysis suggests that their social role is the key to understanding their overall performance. Although family farms cannot fully achieve scale economies, their intense connection with local communities and cultural tradition allows them to give relevant contributions in terms of food security, preservation of traditional ways-of-life, land and environment.

These benefits are often overlooked by agricultural statistic systems. Consequently, policy makers and economists cannot achieve a complete evaluation of the phenomenon. Thus, in this paper, statistical measurement of social role of the family farm is discussed in order to assess the need for additional information.

The analysis starts with a brief outline on family farm statistics worldwide using FAO and World Bank databases, in order to identify current trends (FAO 1997, World Bank 2001a). The emerging scenario is later interpreted according to the current theories on the economic and social contributions of small and family farms. The analysis provides insights for the identification of a possible scenario for their future role in the food system. In conclusion, a brief discussion on how agricultural statistics can capture these developments is proposed.

2. Major trends for family farms

Family farming is the most common way of organizing an agricultural activity worldwide. Although economic literature has provided various definitions of family farms (De Benedictis 1995), for the purpose of this paper, we define family farms as economic and social organizations characterized by a significant overlapping between the family and the farming activity. In a family farm, the labor of the farmer and his/her family is a relevant input (unlike, for example, corporate farming based on hired labor). At the same time, farm income, including self-consumption, is a relevant component of total family income (unlike, for example, hobby farming).
Although the importance and the role of family farms in the economic system depends on the economic development of individual countries, in many cases the rural family is still the backbone of the agricultural system.
The number and the relevance of family farms depend on the economic conditions and social institutions of individual countries. Thus, in this analysis, countries are grouped by per-capita Gross National Income (GNI) classes, according to a classification proposed by World Bank (World Bank 2001). World countries are divided into Low Income Countries, where per capita GNI is lower than US$755; Middle Income Countries, where per capita GNI is between US$ 755 and US$ 9,265; High Income Countries, where per capita GNI is higher than US$ 9,266. Table 1 reports selected data, divided by country class.

Table 1: *Selected Data by Class of Per-Capita GNP*

Countries	Population (Million)	Rural population (1)	% labor force in agriculture	Value Added Agriculture (2)	Agricultural holdings (3)	
					Millions	
	1999	1998	1990	1999	Number	Ha
Low Income	2,417	70	68	27	160	300
Middle Income	2,667	35	28	10	55	1,200
High Income	891	23	5	2	15	1,100
Total	5,975	54	49	4	230	2,600

(1) % of total population
(2) % of Gross Domestic Product
(3) estimate

2.1 Low-Income Countries
The majority of Low-Income Countries (LIC) are located in Sub-Saharan Africa and the Far East. In 1999, their population amounted to 2.4 billion people and their per capita income was US$ 410. In these countries, agriculture is one of the most important economic activities. Approximately 70% of the people are considered "rural population" and 68% of the workforce is occupied in agriculture (the world average is 49%). Agriculture produced approximately 27% of the 1999 Gross Domestic Product (GDP), a percentage similar to industry (30%) and services (43%).

Statistical data report 160 million agricultural holdings and 300 million hectares of farmland with an average of approximately 1.9 hectares per household. In the majority of Low-Income Countries, family farms are the backbone of agriculture. Their number is rising due to the increase in arable land and agrarian reforms. For example, in India, 99.8% of agricultural holdings are represented by family farms with an average extension of 1.5 hectares. In this country, the number of agricultural holdings increased from 61.8 million in 1950 to 97.2 million in 1990, with a remarkable increase in farmland.

2.2 Middle-Income Countries

Middle-Income Countries (MIC) are located prevalently in Latin America, East Europe and Asia. The total population in 1999 was 2.7 billion over a total area of 67 million Km^2. The average per capita income was US$2000, almost five times higher than LIC. In these countries, agriculture is still a relevant economic activity: the agricultural sector produces approximately 10% of the total GDP, a percentage significantly lower than industry (36%) and services (54%). The rural population is 35% of the total.

In 1990, MIC reported 55 million agricultural households and a total arable land of 1.2 billion hectares. The average farm extension was 22 hectares. The general trend showed an increase in the number of farms due to the intense deforestation, especially in Latin America. In Brazil, one of the most representative MIC, the number of farms increased from 4.9 million in 1970 to 5.8 million in 1985, with an increase of 27.9% in arable land. Data showed that these changes brought an increase in the relevance of the largest farms. In 1970, agricultural enterprises with more than 1000 hectares owned 39.5% of the total farmland. In 1985, this percentage rose to 43.8. Family farms, although increased in number, had now a lower proportion of available farmland. The development of the Brazilian farm system increased the relevance of larger, industrialized agricultural enterprises at the expenses of smaller farms. However, the social costs of these changes questioned the social equity of the emerging farm system and raised an intense debate on the need for an agrarian reform (Groppo, 1996).

2.3 High-Income Countries

The group of High-Income Countries (HIC) is composed by Western Europe, the US, Canada, Australia, Japan and New Zealand. The total HIC population amounted to 891 million over a total area of 32 million Km^2. The 1999 per capita income is US$ 25,730: 60 times higher than LIC and 13 times higher than MIC. Agriculture contributed to total GDP by 2%, a low percentage compared to industry (30%) and services (64%). In 1998, rural population was 23% of total, half of the world average percentage.

In 1990, HIC reported 15 million agricultural households and 1,100 million hectares of farmland. The average farm extension of approximately 73 hectares was remarkably higher than MIC and LIC. The general trend showed a decrease in the number of holdings; from 1930 to 1990 the number of farms decreased by approximately 50%.

Japan provides a meaningful example of this trend. According to a 1990 census, the Japanese agricultural system was composed almost exclusively of family farms (only 11,000 holdings are either cooperative or corporate farms) and the average farmland was 1.36 hectares. From 1950 to 1990 the number of farms fell from 6.2 million to 3.8 and the total arable land decreased from 6.2 to 5.2 million hectares. Even if total arable land decreased the average size of Japanese farms increased due to a process of concentration.

2.4 General trends

Data confirm an industrialization process in the farm system of MIC and HIC, implied by the increasing concentration of farmland. On the other hand, in LIC, the number of the small and family farms is rising, due to the increase of arable land. Where the institutional context is favorable, family farming is giving way to larger, industrialized enterprises. The presence of infrastructures, off-farm employment opportunities, adequate social institutions and the availability of financial capital is leading to a reduction of rural population and the concentration of farm activities into fewer, larger and industrialized farms. Thus, in more developed countries, financial and human resources move away from agriculture, in search of more profitable allocations. The food demand is satisfied by fewer and more capital-intensive enterprises which are able to produce a larger quantity of food at a lower cost.
In order to investigate the consequences of this trend, two topics must be addressed. Firstly, production efficiency of family farms must be compared with industrialized enterprises, to identify the impact on total food production. Secondly, the social role of family farms must be assessed to appreciate completely the social consequences of the industrialization process.

3. Economic efficiency of family farms

The comparison of small, family farm productivity with large-scale, industrialized enterprises is a controversial topic. Economic literature analyzes the issue in developed and developing countries separately. The differences in the institutions are so relevant that economists have developed two different approaches.
Using empirical evidence, rural development economists argued that there is an inverse relationship between farm size and land productivity in developing countries (Sen 1962, Lau and Yotopoulos 1971, Berry and Cline 1979, Lund and Hill 1979, Rosset 1999). This notion is widely accepted and it gave justification to agrarian reforms worldwide. The rationale of this argument is that, in large farms, a rent-seeking strategy emerges leading to lower land productivity. This argument is suited particularly to low-income countries where the industrialization process of agriculture was not yet been developed.
Empirical studies associated the higher land productivity of small farms with lower labor productivity, compared with larger enterprises (Byiringiro and Reardon 1996, Deolalikar 1981). Authors indicated greater labor intensity as the reason for the relationship. Also, larger farmers are more likely to use hired labor, non-labor variable inputs and capital to increase productivity and compensate for the lack of family labor.
On the other hand, recent studies report cases where no definitive evidence of difference is found in the economic efficiency of large and small farms. These contributions cautioned against misleading generalization about the existence of an inverse relation between size and land productivity (Adesina and Djato 1996, Townsend, Kirsten and Vink, 1998). Thus,

the comparative production efficiency of small farms in developing countries is still an open debate.

Papers related to developed countries focused on the evaluation of scale-economy in agriculture reaching controversial conclusions. Authors provided empirical evidence of the existence of scale returns in agricultural operations and stressed the consistence of this result with the ongoing concentration process in the US (Batte and Sonka 1985, Cooke and Sundquist 1989, Moschini 1990). On the other hand, a recent survey suggests that, after accounting for management, land quality and other specific factors, small family and part-time farms are at least as efficient as large commercial operations (Peterson 1997). According to the author, the concentration process in the US farm system is more likely to be the result of the high costs for agricultural services, rather than the consequence of scale economies.

In recent years, the perspective of the economic debate has changed, especially in developed countries. The emerging dichotomy is now related to traditional versus industrialized farming. From this point of view, farm size is just one of the diverging factors: the gap is due to capital intensity, input availability, human resources and technology. Thus, the emerging competition is based on skills, financial resources, social and market institutions rather than simply access to land.

Innovation has brought significantly higher yield in industrialized farms due to biotechnology, information technology, chemistry and mechanization. The difference in total productivity between farmers who have skills, financial resources, services and infrastructure to take advantage of these new technologies and the other farmers may be impressive. New technologies and services not only increase average yield, but also reduce production risk. Therefore, a likely consequence of agricultural industrialization is an increase in food productivity.

4. Social role of family farms

A comparison based only on productivity between industrialized enterprises and family farms may be misleading, because the latter perform a broad set of social functions in addition to food production. Thus, a brief overview of some of the most important contributions is now discussed.

4.1 Food security

It is generally agreed that food security is one of the most important performances of an agricultural system. Therefore, the relations between family farms and food security must be considered in an unbiased evaluation. In regard to this, two different arguments are raised. Industrialized agriculture is considered a more stable and less expensive food supply. The progressive reduction of agricultural world prices implies that the global agricultural system is in fact able to deliver commodities at a lower price. The reduction of the total expenditure for food is an important achievement especially for LIC. The higher productivity of industrialized agriculture may be better able to meet the food demand of the increasing world population than traditional family farming.

On the other hand, an agricultural system based on small family farms may be able to deliver food even in presence of high transaction costs in the distribution system. Industrialized agriculture based on production concentration and global trade requires efficient distribution systems. In fact, if transport costs are high enough, local family farming may be a more efficient system to deliver food to rural areas. When the delivery system is not efficient, the location of production becomes as important as cost-effective production. This is because transport costs may become so high as to offset the higher productivity of industrialized enterprises. This issue is particularly sensitive for developing countries, where the amount of transport costs in the distribution system may be significant. The possibility of producing food for self-consumption is one of the causes of the already mentioned large presence in LIC. In the absence of efficient market institutions, self-consumption in rural areas represents a means to meet human basic needs.

In addition, it must be pointed out that many governments consider food self-sufficiency as a strategic goal. This objective increases the relevance of local farming versus food importing.

4.2 Land and environment management

Land and environment preservation is considered one of the major benefits of traditional family farming compared with industrialized agriculture. The issue is particularly relevant in the areas facing land erosion and desertification, where agriculture may be a presidium. Recent studies showed that, in Rwanda, small farms have twice the soil conservation investments and their land is no more eroded than larger farms (Byiringiro and Reardon, 1996). The importance of inter-generation relationships within the family is considered an important incentive to a responsible resource management, at least in absence of income constraints.

Land use is a sensitive issue also for developed countries. In a recent survey, Eurostat reported that farmers hold more than half of the territory in Europe (Eurostat, 1998). Given the environmental impact of their economic activity and their spatial dispersion, European farms are one of the driving forces in the landscape management. The environmental externalities of agricultural production become socially relevant and eco-responsible farming is considered a social benefit.

Family farms are considered to have a better impact on biodiversity than industrialized enterprises. Usually multiproduct, non-specialized enterprises, family farms promote diversity in terms of cropping systems, landscapes and biology. Thus, biodiversity can be considered as one of the most important comparative benefits produced by family farms, with respect to industrialized enterprises.

4.3 Income distribution

The structure of the agricultural system may have a relevant impact on income distribution among rural people. A recent report from the USDA stressed the importance of decentralized land ownership as a form of granting more equitable economic opportunities for people in rural areas (National Commission on Small Farms, 1998). An agricultural system based on family farms gives a large number of people a better access to resources, fosters personal responsibility, individual initiative and, in some cases democracy. In a recent debate, supporters of Brazilian agrarian reform used this thesis to stress the

importance of a proper allocation of land property rights (World Bank 1998). A family farm-based agriculture is considered able to achieve a more equitable income distribution than a system composed of fewer and larger industrialized enterprises.
The ownership of the land allows rural populations to produce income either marketing their products or through self-consumption. In both cases, the better access to resources for rural people leads to a more equitable distribution of wealth.

4.5 Preservation of local traditions and culture

Family farms are considered as keepers of the traditional way-of-life and local culture. In many countries (for example, Europe and Japan) this role is highly considered. In this context, people tend to consider family farms as a link with social roots of traditional culture. Thus, family farms are a reserve of social diversity even in the age of globalization. Family farms are also considered the keeper of local food-cultures (i.e. the traditional way of preparing and consuming food), a function that cannot be performed by industrialized agriculture. In fact, although industrialized food systems are introducing "ethnic differentiation" such as Mexican or Italian food, the global market seems unable to preserve the great diversity of food cultures.

4.6 Food quality

The production of a constant supply of quality food is considered another important role of family farms, especially in developed countries. As soon as the nutritional needs are met, consumer demand starts moving from food quantity to food quality: lifestyle, diversity, taste and personal satisfaction becomes consumers' most desired attributes. Family farmers are considered more suited to deliver these attributes. Due to their closeness to final consumers in a specific area, local producers have more information about the desired attributes and are able to deliver them more effectively. Moreover, the connection with local traditions enhances the family farms' capability of differentiating their supply and of providing unique value-added attributes for local consumers.

5. Perspective for family farms

This brief outline suggests that family farms perform a broad set of social functions in addition to food production both in developing and developed countries. Environment management, biodiversity, income distribution, food security, food quality and social diversity are just a few of the joint products produced by family farms. Although the specific contributions depend on the unique characteristic of each country, family farms are able to perceive the demand of the local community. Being part of the local community, family farms seem to be able to adapt to their special needs.
Although the competition with industrialized enterprises will be increasingly challenging, the social importance of family farms makes the extinction hypothesis unlikely. Family farms are expected to strengthen their unique characteristics and differentiate their supply moving from commodity production to value-added agriculture, in a multi-functional approach. The core of the multi-functional approach is to translate the social benefits of

family farming into economic value, through differentiation, eco-responsibility, quality and a special deal with consumers and governments. A multi-functional farm, being able to create economic value by providing additional services to consumers and community in addition to food production, can compete in the global market by building a differentiation strategy based on its own strengths. Thus, in the near future, the multiple benefits produced by family farms are expected to become increasingly relevant, ensuring survival and competitiveness with industrialized agriculture both in developed and developing countries.
However, family farms are expected to perform different functions, according to the economic development of their community. In LIC and MIC, their most important roles are undoubtedly related to food security and income distribution. The relevance of these functions may explain their increasing presence. However, it must be stressed that the different institutional contexts are leading to different scenarios in LIC and MIC. In the former, family farms are the dominant forms of agricultural organization and their number is increasing together with population and arable land. In the latter, the presence of infrastructures and marketing channels allows industrialized agriculture to develop. Thus, we have coexistence of both organizational models. Family farms and industrialized enterprises compete in the same markets, with different strategies and with different social functions. Industrialized farms are involved in large-scale commodity production and export, while family farms are focused on meeting the demand of local markets and food security in rural areas. In this context, family farms cannot differentiate their products. Because of the low per capita income in these countries, the development of a multi-functional approach to agriculture for family farm is unlikely in the short term. In fact, the income constraint makes the majority of consumers unwilling to pay a price premium for eco-responsible products and quality food. The concentration process of farmland suggests that the relevance of industrialized agriculture is increasing at the expenses of smaller farms. In the absence of public intervention, this trend could lead to serious consequences in terms of income distribution in rural areas.
In developed countries, the scenario is different. The presence of efficient distribution channels and the high per capita income reduce the relevance of food security and income distribution functions. Being challenged on their local markets, non-industrialized family farms either evolve or are forced out of business. Thus, biodiversity, landscaping, environment management and food quality are becoming the most relevant social functions and the multi-functional approach a viable and sensible strategy.

6. Family farms and agricultural statistics systems

A basic condition for the success of the multi-functional approach is the full recognition of family farm social role by consumers, communities and governments. From this point of view, agricultural statistics have the important role of providing complete information to public opinion and decision-makers. Agricultural statistics overlooking the social role of family farms could bias the general appreciation of their contribution preventing decision-makers from performing complete and correct cost-benefit evaluations.

Traditionally, in agricultural statistics, the focus is on accuracy in the estimates of food production, due to food safety concerns and the needs of data for economic planning. However, in order to provide complete information, agricultural statistics should investigate the relationship between farms and their economic, social and ecological environment. The contribution to food security in marginal areas, the impact on income distribution in rural communities and the function of land management and environment preservation are relevant factors that should be carefully evaluated to avoid significant judgement bias.

Table 2 summarizes and categorizes the data collected in a small sample of agricultural census. The survey considered 17 selected agricultural census questionnaires from the FAO statistics library. The collected data can be summarized in the following groups:

1. Demographic data (for example: age and sex of the farmer and his/her family). These data have been collected by all the questionnaires.
2. Land use. Data related to the presence of woodland and wetlands, crop acreage, buildings, etc. have been collected by all the censuses, representing a sort of common ground across the countries.
3. Livestock. Almost all the censuses (88.2%) collected data about the number of animals owned by the farm, by species.
4. Agricultural employment. 88.2% of the questionnaires collected data about family and hired labor. The number of workforce and the number of working days are the most common statistics within this group.
5. Information about equipment and machinery is collected by 76.5% of the selected censuses.
6. Farm practices. This group reports data about crop rotation, use of chemicals and fertilizers, irrigation, etc. It was present in 70.6% of the selected censuses and represent the most common attempt to evaluate some aspects of the environment impact of agriculture.
7. Production statistics. 52.9% of the censuses collect information about agricultural production, such as yields, dairy production, etc.
8. Other data are related to marketing (41.2%), off-farm activities (29.4%), finance (23.5%), policy evaluation (11.8%), food quality (11.8%) and on-farm food processing activities (5.9%). Often, these data are collected to meet specific information needs of individual countries.

The analysis showed that the majority of the censuses focused on demographic, labor and structural statistics (such as farmland, commodity production, equipment and machinery, etc.). Some questionnaires asked for data related to farm finance, marketing and family labor. FAO recommendations from the programme for the world census of agriculture 2000 (WCA 2000) were met by the majority of the selected countries, although to a different degree of fulfilment.

Table 2 shows the differences among the census questionnaires. LIC censuses are more focused on structural statistic (i.e. demographic, land use, employment and equipment). They overlook farm practices and, in 50% of the cases, production data. The other groups (marketing, off-farm activities, finance, etc.) are ignored as well.

MIC questionnaires, in addition to the data reported by LIC, collect data about farm practices and marketing. Information about off-farm activities, food quality, food processing and policy are collected in few HIC only.

Table 2: Data Collected by 17 Agricultural Censuses

			Categories of Data Collected												
Country	Year	WB	Demo-graphic	Land Use	Live-stock	Employ-ment	Equip-ment	Farm Practices	Produc-tion	Marketing	Off Farm Activities	Finance	Policy Evaluat.	Food Quality	Food Proces.
Congo	1988/89	LIC	✓	✓	✓	✓	✓		✓	✓					
Senegal	1997	LIC	✓	✓	✓	✓			✓						
Leshoto	1989/90	LIC	✓	✓	✓	✓	✓		✓		✓				
India	1985/86	LIC	✓	✓			✓								
Pakistan	1990	LIC	✓	✓	✓	✓	✓	✓				✓			
Bangladesh	1996	LIC	✓	✓		✓	✓								
Mexico	1991	MIC	✓	✓	✓	✓	✓	✓	✓	✓					
Argentina	1988	MIC	✓	✓	✓	✓		✓				✓			
Chile	1996/97	MIC	✓	✓	✓	✓	✓	✓	✓	✓					
China	1996/97	MIC	✓	✓	✓			✓			✓				
Canada	1991	HIC	✓	✓	✓	✓	✓	✓		✓		✓			
USA	1997	HIC	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
Slovenia	1991	HIC	✓	✓	✓	✓		✓			✓				
France	2000	HIC	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	
UK (*)	1997	HIC	✓	✓	✓	✓	✓	✓	✓						
Italy	2000	HIC	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓
FAO(**)			✓	✓	✓	✓	✓	✓	✓						

(*) Aggregation of two questionnaires: Scotland plus England and Wales
(**) Recommendations from FAO programme for the world census of agriculture 2000 (WCA 2000)

In general, questionnaires of more developed countries collect more information than others. This difference is explained not only by the amount of the financial investment in agricultural censuses but also by the different functions performed by agriculture. In HIC, aspects related with food quality and off-farm activities are considered more relevant; thus, they obtain greater attention. However, it must be pointed out that, even in HIC, the social role of family farms is not fully appreciated, because the focus of the census is on structural and production data.
In table 3, a classification of data present and missing in the majority of agricultural censuses is proposed.

Table 3: *Agricultural Censuses*

What (usually) can be found:	What is missing:
Land use and property rights	Food security
Cultures and livestock	Food quality
Equipment and machinery	Environment protection
Marketing	Tradition and local culture
Finance	Income distribution
Farm Practices	Inter-generation relationships
Demographic	Biodiversity
Labor statistics	

Table 3 suggests that agricultural censuses are missing relevant aspects of family farm social role. However, the statistical measurement of social and environment impact is a difficult task due to the complexity of the subjects. Differently from physical outputs and land use, often, social benefits are not clearly defined notions. Thus, the design of specific indicators is still a work in progress, although some results have been achieved in the environmental management field (Commission of the European Communities 2000).

7. Summary and conclusions.

The globalization and industrialization processes of world agriculture are posing a serious threat to small family farmers. WTO negotiations, fostering free trade, will probably lead to an increase in competition and to a reduction in public support to agriculture. In this scenario, many farmers are exposed to the risk of being driven away from the food market.
The potential crisis of the family farm system may have serious consequences. Performing a broad set of social functions in addition to food production, family farms have a relevant role in local communities and in many economic systems. Thus, their future is of public interest.
The extinction of family farm is unlikely in the near future because of their large number worldwide, their economic relevance in developing countries and their social role in rural areas. However, radical changes are expected. To meet the challenges family farms need new strategies.

Family farms are expected to evolve either toward industrialized agriculture or to a multi-functional approach, able to give economic value to their unique characteristics. The overall direction of the change depends on the specific characteristics of individual farm systems.
Currently, agricultural statistics are not designed to capture the evolution of farm systems. In particular, the social role of agriculture is often overlooked and the total contribution of family farms is undervalued. The analysis of the potential role of family farms suggests that production and structural statistics may not be sufficient to evaluate completely their contribution to society. Family farm externalities may be as relevant for society as food production. Thus, many agricultural censuses may be unable to provide a complete evaluation of the social role underestimating their contribution.
In order to overcome this problem, agricultural statistic censuses should collect systematic data about family farms multiple benefits. In particular, data related to food security and quality, social aspects and environment impact should be collected. The analysis of the social role of agriculture requires a broad set of new indicator of the impact of agricultural activity on local communities and environment.

References

Adesina A. and K. Djato (1996): Farm Size, Relative Efficiency and Agrarian Policy in Cote d'Ivoire: Profit Function Analysis of Rice Farms, *Agricultural Economics* 14, 93-102

Batte M. T. and Sonka S. T (1985).: Before and After Taxes Size Economies: an Example for Cash Grain Production in Illinois, *American Journal of Agricultural Economics* 67, 600-609

Berry R.A., Cline W.R. (1979): *Agrarian Structure and Productivity in Developing Countries*. John Hopkins University Press, USA

Byiringiro F. and Reardon T. (1996): Farm Productivity in Rwanda: Effect of Farm Size, Erosion and Soil Conservation Investments, *Agricultural Economics* 15, 127-136

Commission of the European Communities (2000): *Indicators for the integration of environmental concerns into the Common Agricultural Policy.* Communication from the Commission to the Council and the European Parliament. Brussels.

Cooke S. C. and Sundquist W. B. (1989): Cost Efficiency in US Corn Production *American Journal of Agricultural Economics* 71, 1003-1010

De Benedictis, M. (1995): *Agricoltura Familiare in Transizione* INEA, Rome, Italy

Deolalikar A.B. (1981): The Inverse Relationship between Productivity and Farm Size: a Test Using Regional Data from India *American Journal of Agricultural Economics,* May, 275-279.

Eurostat (1998): *European Landscapes: Farmers Maintains More Than Half of the Territory.* Statistics in Focus: Agriculture.

F.A.O. (1995): *Programme for the World Census of Agriculture 2000.* Rome

F.A.O. (1997): *Report on the 1990 World Census of Agriculture*. Rome

F.A.O. (2001): *Food Outlook*, Rome

Groppo P. (1996): Agrarian Reform and Land Settlement Policy in Brazil: Historical Background, *SDdimensions* June 1996 on www.fao.org

Lau L.J. and Yotopoulos P.A. (1971): A Test for Relative Efficiency and Application to Indian Agriculture, *American Economic Review* 61, 94-109

Lund P.J., Hill P.G. (1979): Farm Size, Efficiency and Economies of Size, *Journal of Agricultural Economics* 30, 145-157

Moschini G. (1990): Nonparametric and Semi Parametric Estimation: an Analysis of Multiproduct Returns to Scale *American Journal of Agricultural Economics* 72, 589-596

National Commission on Small Farms- USDA (1998): *A Time to Act* Washington D.C.

Peterson W. (1997): *Are Large Farms More Efficent?* Univesity of Minnesota, Department of Applied Economics, Staff Paper P97-2

Rosset P. (1999): *The Multiple Functions and Benefits of Small Farms Agriculture* Policy brief prepared for "Cultivating Our Futures" the FAO/Netherlands Conference on the Multifunctional Character of Agriculture and Land, 12-17 September 1999, Maastricht, NL

Sen A.K. (1962); "An Aspect of Indian Agriculture" *Economic Weekly* Feb. 243-246

Townsend R.F., Kirsten J., Vink N. (1998): Farm Size, Productivity and Returns to Scale in Agriculture Revisited: a Case Study of Wine Producers in South Africa, *Agricultural Economics* 19, 175-180

World Bank (1998): *Report on Brazil-Land Reform and Poverty Alleviation Project.* Project report 7 December 1998

World Bank (2000): *World Development Indicators*. Washington

World Bank (2001a): *World Development Report 2000/2001*. Washington

World Bank (2001b): *Classification of Economies* Internet site www.worldbank.org

Economic and Social Consequences of Biotechnology: A Scenario Analysis

Dave D. Weatherspoon
Rm. 213D Ag. Hall, Michigan State University, East Lansing, MI 48824, USA
e-mail: weathe42@msu.edu

H. Christopher Peterson
Rm. 80 Ag. Hall, Michigan State University, East Lansing, MI 48824, USA
e-mail: peters17@msu.edu

David Neven
108 Cook Hall, Michigan State University, East Lansing, MI 48824, USA
e-mail: nevendav@msu.edu

1. Introduction

Over the years agricultural technology has created remarkable commodity production growth rates and enhanced general economic growth through food production, manufactured goods and trade for most nations. Biotechnology holds the promise of continuing this remarkable record. There is a long list of potential benefits of biotechnology but unfortunately the perceived costs/risks are also many. These concerns have lead to significant consumer reluctance to accept the technology and, in some cases, outright consumer rejection of the technology. To discuss the future of biotechnology, scenario analysis is used to examine the social and economic impact of biotechnology on industrialized and emerging nations. Four scenarios are discussed in detail: biotechnology may be formally or informally banned (Scenario 1), fully accepted (Scenario 2), marketed through strict labeling (Scenario 3), or limited to non-food applications (Scenario 4). Consumer acceptance of this technology will be the key to determining which scenario becomes the future for each nation. The likelihood of each scenario is different for each nation, the U.S. will most likely evolve into scenario 2 or 3, while in the EU scenarios 1 or 4 are more likely. Determining the future for emerging nations is extremely complex and dependent on several factors like malnutrition rates, environmental safety and historical trading routes. Each scenario has a major impact on small producers worldwide which ultimately influences the health of rural communities. The analysis indicates that emerging nations are the most sensitive to the timing of decisions being made about the future of biotechnology. If biotechnology becomes a reality, new data will be required to assess the social and economic impact of this technology.

2. Economic and Social Consequences of Biotechnology: A Scenario Analysis

Over the years agricultural technology has created remarkable commodity production growth rates and enhanced general economic growth through food production, manufactured goods and trade for most nations. As yet another wave of technological innovation, biotechnology holds the promise of continuing this remarkable record. There is a long list of potential benefits of biotechnology which include, but is not limited to, increased production through reducing impacts of pests, improved storage, enhanced health effects from potential stacked traits that fight resistance and increase nutritional value, and improved environment through reduced pesticide use. Unfortunately, the potential and perceived risks are many as well, including threats to environmental and health safety, the emergence of poor nations as inappropriate technological testing grounds, and the potential for other unintended (and as yet unknown) negative consequences of a dramatically new technology. Collectively, these concerns have lead to significant consumer reluctance to accept the technology and, in some cases, outright consumer rejection of the technology. Less commonly discussed is that biotechnology may have other impacts of concern on the social aspects of food production, historical trade routes, economical size of farms and possibly the number of producers.

To analyze these controversial issues about biotechnology, the paper begins with a description of three key uncertainties-food security, environmental/health impacts, and consumer reaction-that will define the future for biotechnology's use in food applications. Based on these uncertainties, four alternative future scenarios for biotechnology are presented. The probabilities of these various scenarios developing in reality are then examined in the context of both industrialized and emerging nations. The likelihood of the various scenarios, i.e., the likelihood of biotechnology's acceptance or rejection, is dramatically different based on what type of nation is considered. Distributional impacts of these possible futures are then presented with a focus on emerging nations. Finally, the data requirements are defined for properly monitoring the resolution of the uncertainties and the analysis of distributional impacts.

What evolves from the analysis is a complex image of biotechnology's future. For emerging nations especially, the technology offers both immense benefits and extreme risks. Emerging nations may be torn between feeding their populations at home and losing trading partners abroad. Timing may be everything. If the emerging nations wait too long to adopt the technology and it proves beneficial, they may miss income opportunities through trade and may end up importing food products that they should actually be producing and exporting. If they adopt too quickly and the technology proves to have net costs instead of benefits, they may lose both at home and abroad.

2.1 Fundamental Uncertainties about Biotechnology's Future

The controversy over biotechnology is driven by a number of key uncertainties. The first uncertainty is whether the growing population of the world can be fed without the emergence of a new "green revolution" Biotechnology has been argued to be this needed new technology. Some population growth projections would clearly argue that this new revolution is needed. Even so, biotechnology may or may not be the new technology needed. However, the slowing of world population growth rates, the emergence of expanding commodity supplies from emerging nations, and the current glut of food commodities worldwide would be counter arguments against this need. Thus it is uncertain whether biotechnology is needed to assure food security.

The second uncertainty is whether biotechnology results in net environmental and health benefits or costs. On the benefits side, biotechnology reduces reliance on agri-chemicals in the food chain and holds open the promise of new beneficial health attributes being engineered into food, e.g., golden rice. On the cost side are the concerns over superweeds, superbugs, loss of beneficial insects or plants, terminator technology, allergic reactions, loss of nutrition, and other unintended environmental and health consequences of a new technology. The argument has been advanced by some that these costs have been overestimated and that "good science" can prove this. Again, it is unclear whether biotechnology will bring about net gains to the environment or human health.

The third uncertainty relevant to the future of biotechnology is the public/consumer reaction to the technology. In reality the first two uncertainties will almost assuredly not be resolved perfectly in one direction or the other, i.e., absolutely needed/absolutely safe or totally unneeded/totally unsafe, the reaction of the public to the likely tradeoff decisions related to biotechnology will have a major impact on the nature of it's adoption. This has already proven to be true. There will likely be perceived differences across the public about the costs and benefits related to either food security or environmental/health impacts. As a result, one part of this uncertainty is whether individuals in a market setting will be allowed to respond individually or instead governments will respond en masse on behalf of the public in the form of broad regulation. Another part of this uncertainty is whether the consumers in the industrialized versus emerging nations may respond differently in how benefits and costs are assessed and traded off.

By their very nature, these three uncertainties cannot be resolved in the short-run. Only the emerging actual changes in population growth, biotechnology impacts on environment and health, and public/consumer reaction can resolve the uncertainties. Unfortunately, decision makers both private and public must make decisions today about whether to continue, expand, or abandon investments in the technology. Without such decisions, the technology may not be available if needed and may not be appropriately managed to mitigate adverse impacts if they emerge.

2.2 Four Scenarios for Biotechnology's Future

Based on the techniques of scenario analysis (Schoemaker, 1995), four possible futures could emerge based on these uncertainties:

> **Scenario 1: Biotechnology is banned, either formally or informally.** The most direct road to this scenario is that the food security needs do not emerge, but the adverse environmental/health impacts do. The public reaction will be clear and negative in this case, either individually or en masse. A less direct road to this scenario could also emerge. The perceived adverse impacts of biotechnology come to be viewed by the public as a set of risks that are not seen as outweighed by any level of potential benefits. Biotechnology may thus be banned formally by government or informally by consumer

choice in the marketplace. The U.S. nuclear experience provides evidence of this less direct path. Although no formal ban exists, public reaction has in effect lead to no new nuclear power facilities being built in the last 20 years.

Scenario 2: Biotechnology becomes fully accepted in the marketplace. Consumer-oriented biotechnology products become available, and biotechnology follows the pattern of other past agri-food innovations. Again, there are both direct and indirect paths to this scenario. The direct path is that the food security needs do emerge, and biotechnology proves to be safe in meeting these needs. The public reaction will be clear and positive in this case, either individually or en masse. The indirect path is that the food security need emerges, safety protocols (public and/or private) are developed to allay public concerns, and biotechnology emerges in the court of public opinion as the safest of available alternatives. Either path would result in biotechnology ultimately becoming fully accepted even though the indirect path would have a longer period of controversy before consensus emerged.

Scenario 3: The food system adopts a "three labels" approach to biotechnology. Given the nature of the food security and environmental/health uncertainties, a solution that may evolve and has been adopted by some already, e.g., Japan, is to label all food by its method of production-conventional, biotechnology, and organic. Individual consumers thus choose in the marketplace based on their individual weighing of the risks and benefits. The public reaction is thus allowed to be resolved on an individual basis rather than en masse.

Scenario 4: Biotechnology is used only in non-food applications. Consumers are unwilling to have biotechnology used in food, but would allow commercial or industrial applications, e.g., pharmaceuticals, replacement of industrial chemicals with biologically based alternatives. Based on the recent StarLink experience and other needs to allay public concerns, a system of careful control of biotechnology crops and livestock emerges. Such agricultural products are produced in environmentally-controlled, manufacturing-like facilities. Strict system segregation from food uses is maintained to the public's satisfaction.

The scenario analysis suggests that the third uncertainty-consumer reaction-is critical to determining which future scenario actually occurs. The first two uncertainties will only be resolved over an extended time while the third one can be and is already being played out in the near term. All scenarios are sensitive to timing for all countries but timing is more crucial for emerging nations. The risk of investing today with no guarantee of a future market (domestic or international) would be devastating for emerging nations as would the cost of resolving adverse environmental or health impacts, should they develop.

The resolution of the third uncertainty-public/consumer reaction-depends on the completeness and quality of the information that the consumer has for decision making. In essence, the consumer is faced with a classic information problem akin to the market for lemons (Akerlof, 1990). The consumer is being asked to purchase a product whose attributes cannot be known at the time of purchase, i.e., the true costs and benefits of biotechnology. The consumer must thus fear opportunism on the part of the seller, i.e., the life sciences industry, who is presumed to have superior private information. The seller has incentives to hide information if it is adverse to the buyer's interests and to share it if in concert with the buyer's interests. However, the buyer cannot know a priori which is true of the information-sharing practices of the seller. The history of the tobacco industry is a case well established in the minds of the consumers as a perfect example of this information problem.

The problem for the consumer is further complicated by the fact that there are many potential "sellers" of information about biotechnology. Governments, environmental groups, the scientific community, non-governmental organizations (NGOs), and food industry firms outside the biotechnology industry also have interests in consumerss decisions about this issue. The consumer must gauge the integrity of each of these information sources.

The critical question relevant to public/consumer reaction becomes: Who has the integrity to interpret the risks and benefits of biotechnology to the general public and the food consumer? Given the information problem posited, the integrity of the messenger will be a key driver of the acceptance or lack of acceptance of biotechnology and thus of the future scenario that emerges. (For an expanded discussion on integrity see Peterson and Weatherspoon, 2001)

2.3 Applying Scenarios to Emerging and Industrialized Countries

The future scenario that emerges for biotechnology will depend on consumers' evaluation and ultimate response to the integrity of relevant messengers about technology and the existing risk factors within their respective countries (no major food issues to malnutrition, human, plant and animal diseases, national security, etc.). Consider first the likely differences in the scenarios that may emerge for industrialized nations, such as, North America versus Europe.[1]

In North America (particularly within the NAFTA countries), the probability that either Scenario 2 (full acceptance) or Scenario 3 (consumer choice through labeling) will emerge is high. The US has a long history of successful agri-food system monitors with high levels of both perceived and real integrity, e.g., USDA, FDA, EPA, and university scientists. The USDA, FDA, and EPA have given approvals to biotechnology. Canada, much more than Mexico, has a similar system of monitors to the US, and both Canada and Mexico closely follow the lead of the US on these types of issues.

The probability of biotechnology playing a major role in the future of the EU is much lower than in North America. Scenarios 1 (banning of biotechnology) or 4 (limited use to non-food applications) are far more likely to emerge. The EU's government monitors have lost perceived integrity recently due to the way they handled the outbreaks of mad cow disease and foot-and-mouth disease. The highly promotional manner in which biotechnology was introduced in Europe by the life sciences companies also limited the life science companies perceived integrity. Contrarily, the NGOs have vigorously protested biotechnology and have appeared to have influenced policy which has limited the consumption and production of biotech products in the EU. The current movement in Europe to increase food safety will presumably be

[1]Hill and Battle (2000) provide a useful analysis of the EU-US GMO debate.

expanded to the biotechnology area. For example, COLEACP (a liaison committee for EU-ACP horticultural trade) has just received a large contract from the EU to work directly with exporters from emerging nations to make sure that their products have an acceptable level of pesticides. This effort includes training and labs to test in country which will result in transparency and liability throughout the system. This effort can and most likely will be expanded to test for various biotech enhancements in food products (Guichard, 2000). Unless risks are reduced and products are put in the market channel that address consumer benefits (cure diseases, bio factories with positive environmental impacts, etc.), the prospect of biotechnology food products being successful in the EU is low. The other option is that the governing bodies of the EU food system regain their influence and decide that biotechnology is critical for the future.

Arguments have been advanced that the development of consumer benefits from biotechnology (as opposed to the agronomic benefits created to date) will in and of itself create consumer acceptance. In effect, this is an argument that the consumer will forget the potential costs and risks in the face of enhanced benefits. Now that the public and food consumers are aware of the issues related to biotechnology, it is not very likely that the emergence of new benefits will eliminate the controversy. The balance of benefits and costs may improve in favor of benefits, but consumers will remain skeptical. For the consumers of the industrialized world most especially, the new benefits may have to be extremely valuable to counterbalance the uneasiness over potential risks.

As a second case, consider the differences of reaction across the emerging nations. The future of biotechnology is more complex for these nations. The economies of many of these nations are dependent on the US, EU or Japan for trade. Therefore, local decisions about biotechnology may be heavily influenced by the final market for the product. In sub-Saharan Africa where many nations have historic trading ties with the EU, the leaders in most of these countries would like to reduce their malnutrition rates and increase income from agricultural trade. Such leaders are torn between the potential benefits of biotechnology and the demands of products moving from home markets to the EU. The economic messages remain mixed as to what to do with the technology. Thus far, only South Africa and Kenya have ongoing trials of biotech products in sub-Saharan Africa. Any of the four scenarios for biotechnology could thus emerge in sub-Saharan Africa given the mixed messages that exist there. In contrast, several South American countries have already commercially produced biotech products and are conducting experiments on new products. The export markets for these nations are largely North American based which is arguably pro-biotech. Hence, the South Americans are taking less of a risk relatively than the South Africans and Kenyans who primarily export to the EU.

Predicting the scenarios for the emerging nations is further complicated by the existence of a centrally-planned economy or an authoritarian regime where government is the sole voice on the biotechnology issue. China is a good example where the government has decided that biotechnology is the key to feeding their population. Consumer choice to avoid biotechnology products may not exist in this system. Scenarios 2 (full acceptance) is created by fiat, but may unravel to Scenarios 1 (banning the technology) or 4 (limiting to non-food uses) if subsequent events prove the liabilities rather than the benefits of the technology. In any event, the surprising result is that Chinese consumers are in effect facing the same situation as North

American consumers currently. Neither can predetermine the presence of biotechnology in their food and thus neither has a real choice. North American consumers do have the ability to create Scenario 3 (private choice through labeling) while the Chinese consumers are not likely to have this choice except perhaps when such choice serves the demands of valued export markets.
A third sub-case for the emerging nations exists and has the highest probability in the poorest of the poor countries. In this system there are several key breakdowns in society that may lead to business representatives having the greatest influence on biotechnology's adoption. These key breakdowns include one or more of the following: dysfunctional governments, extremely high food security risk, a fragile environment, and/or abundant health risk. The perceived risks of biotechnology may be lower than the real risks from these breakdowns. A biotechnology company could provide great opportunities to producers, traders and the rural community. An anecdotal example of this comes from South Africa which does not represent the poorest of the poor as a country, however, it is a nation best described as both first and third world. Monsanto created a pilot program (subsidized the cost of buying the seed) for resource poor producers so that they would try Monsanto's BT Cotton (Brink). These producers were reported to have increased their annual profits by $150 on average. Technology adoption becomes driven by real benefits in the face of very high risk. If this example is expanded to a truly destitute country, it is a possibility that a private business, e.g., Monsanto, could become the sole driver of adoption. Presently, most biotechnology firms do not see ample returns on their investment and thus bypass these nations.

2.4 Distributional Impacts of Possible Futures

There are distributional impacts that vary across scenarios. These impacts are more sensitive and critical for emerging nations and deserve special treatment in this section. Focus is placed on small producers and the general economic and social health in these nations. What is not commonly discussed is that biotechnology will strongly influence the social aspects of food production, trade competitiveness, economical size of farms, and the number of producers. For the discussion that follows, scenarios 2 and 3 are considered together because they are relatively similar in regard to distributional impacts while 1 and 4 are considered together because of their similarity to each other and their distinct differences with 2 and 3.

<u>Scenarios 2 & 3</u>

The distributional impacts of scenarios 2 and 3 (biotechnology is fully accepted or there is a three label system) can be both specific to the situation of a particular country and crosscutting for small producers everywhere in the emerging nations. Consider the small producer impacts first. Under either scenario, commodity production will be heavily influenced by biotechnology's presence. Biotechnology R&D has been focused primarily on the largest volume agronomic crops by the large life sciences firms. This trend is expected to continue. The impact this may have on smaller producers worldwide is that they will likely not be able to compete in these markets because this technology is not scale neutral. Small producers may be at a disadvantage for several reasons:

- the specialty crops focused on by small producers have not been the focus of biotechnology research and commercialization. Only universities and government-sponsored research labs usually conduct research on specialty crops;
- biotechnology enhanced inputs are expensive to use and require more working capital;
- although the technology is embedded in seed, the full utilization of the technology requires sophisticated management and marketing practices that create a significant learning and investment curve for small producers;
- marketing options may be limited depending on which scenario evolves and where.

All of these factors indicate that small producers may be exposed to especially high risk if they adopt this technology.

Although Scenario 2 (biotechnology is fully accepted), will likely have a negative impact on small producers of commodity type products, small producers may have a competitive advantage in producing biotech products that are labor intensive or that are unique in their production methods or preservation of identity throughout the system. Products that have medicinal properties would be good examples of where small producers may play a major role. Timing becomes crucial for emerging nations if scenario 2 becomes reality. The poorest emerging nations need to be in the first wave of adopters and invest in biotechnology to be dominant in the world market and to manipulate the technology to address their local needs (particular pest type, disease resistance, drought resistance, fortification with certain vitamins, etc.). If emerging nations wait too long, they may miss income opportunities through trade and may ultimately import the products that could have been grown locally. Several South American countries are at the cutting edge of this technology and are competitive in commodity markets. Kenya has conducted trials with biotech crops but has not yet commercially produced anything for the market. South Africa has commercially produced corn and cotton, but it is not clear if those products were exported or consumed domestically. If biotechnology is fully accepted, these emerging nations will have an advantage over other emerging nations since the region specific R&D will have been completed by the time the technology is fully accepted by consumers.

Scenario 3, three label system, does allow small producers to be competitive in all three categories: organic, traditionally produced crops and biotech products that are not commodities as discussed under scenario 2. The negative impact this approach could have is that all of the products will have to be tested to make sure that they meet the standards outlined for each category. This is another step that small producers are ill prepared to make. The extra costs of this certification stage may prohibit small producers in poor nations from participating in the formal market.

The positive aspect of scenarios 2 & 3 is that a powerful new technology will be available to address the malnutrition needs in emerging nations. Fortified products and medicinal foods may have the ultimate positive impact on poor rural communities worldwide. Having healthier workers alone may be worth the investment for poor nations. The potential negatives are the social impacts and the environmental and health safety issues. Socially, biotechnology may reduce the number of viable producers in the rural areas and create a large number of displaced people. The rural to urban migration in many emerging nations have overwhelmed the available resources. In sub-Saharan Africa, the economic ramifications are tremendous given that 70-80 percent of the population lives in rural areas and are connected with agriculture. Safety risks associated with the introduction of biotechnology must also be monitored. Who will bear the cost of monitoring, especially in emerging nations? The potential for emerging nations to be the testing ground for some of these new technologies is high. The need for an international monitoring agency may evolve under these scenarios because the individual emerging nations do not have the capital to invest in monitoring at the local level.

Scenarios 1 & 4

The distributional impacts of scenarios 1 and 4 (biotech is banned or limited to non-food applications) are dramatically different from 2 and 3. If the poorest emerging nations were in the first wave of adopters of this technology, this would irreparably harm their producers, particularly the small producers. Producing this technology without a clear indication that there will be a market is extremely risky for these nations. In addition, they are risking all agriculturally related export markets because the world may perceive them as contaminated. Therefore, a nation's timing on when or if to allow biotechnology in the country is crucial.

There are no anticipated positive income impacts on small producers since they will not have access to this cutting edge technology in their production practices. Hence all returns associated with biotechnology for scenario 4 go to the inventors and the special firms that can produce products in a bio-secure environment. At the same time there are no additional negative impacts on small producers or rural areas under either of these scenarios with the exception of those regions that invested early in this technology.

The negative impact of these scenarios is that a tool to address malnutrition needs will not be available in the short run. The positive aspect is that the perceived environmental and health risks associated with this technology are no longer a concern.

2.5 Data Needed to Document Impact of Biotechnology on Rural Areas

Given the significant potential benefits and costs of biotechnology, policy makers who want to influence biotechnology's adoption and impacts need to have appropriate data to support decision making and monitoring. Two distinct types of data are needed - data to track the resolution of the uncertainties surrounding the technology and data to document the distributional impacts.

The uncertainties defined earlier – food security needs, biotechnology benefits and costs, and consumer reaction - each demand somewhat different data to monitor how the uncertainty gets resolved over time. The need for biotechnology to contribute to food security will depend upon trends in population growth, demographics, commodity and food production capacity, import/export conditions, and the emergence of non-biotechnology alternatives to improve yields and overall production. To the extent that population pressures ease and alternative production develops, biotechnology becomes more expendable. Conversely, the more population pressures mount and alternatives fail to materialize, biotechnology becomes more necessary whatever the costs. Biotechnology's ratio of benefits to costs will depend most heavily upon whether the downside risks to environment and health actually emerge. Therefore, monitoring of a variety of environmental and health impacts becomes critical, including (but not limited to) changes in beneficial and pest insect populations, biotechnology-induced changes in quantity, quality, and productivity of crops and livestock, changes in bio-diversity, and changes in health among consuming groups. Monitoring the continuing scientific research on the benefits and costs of biotechnology will also be needed. Finally, tracking consumer reaction to biotechnology-based products becomes critical to making marketing, production, and distribution decisions.

All of these data are equally relevant to industrialized and emerging nations. Unfortunately, the emerging nations will have far less capacity to create and monitor the necessary information. From a practical perspective, they may become dependent on interpolating information from experiences in the industrialized world to support their own local decision making. This dependence will create concerns in its own right. One option to alleviate this concern would be to develop an international monitoring agency so that standards are consistent worldwide and the costs of such a system are more broadly shared.

To monitor the distributional impacts of biotechnology, data focused specifically on rural areas of the emerging nations will be critical. Such data would include (but not be limited to): (1) changes in rural population size and distribution, (2) economic activity, including agriculturally related value-added activities, (3) size of an economically viable farm, including minimum effective scale per crop, (4) types of crops produced by method of production (biotechnology enhanced, traditionally produced, and organically produced), location, and size of operation, (5) percent of product actually sold versus personal consumption, (6) percent consumed domestically versus sold globally, (7) commodity and product flows by market channel into formal or informal markets, and (8) tracking of rural incomes per capita. Base-line as well as time series data would be needed to accomplish the monitoring task. Again, this information system would impose heavy costs on emerging nations unless international agencies step forward to help.

The information base required to monitor either the path of the uncertainties or the distributional impacts would need to be carefully studied to assure that the appropriate data is collected and the appropriate analysis techniques are used for decision making.

3. Summary and Conclusion

Three key uncertainties - the real need for biotechnology to assure food security, biotechnology's real balance of benefits and costs, and public/consumer acceptance of the technology-give rise to four scenarios for biotechnology's future. The technology may be formally or informally banned (Scenario 1), become fully accepted (Scenario 2), move to market through strict labeling (Scenario 3), or limited to non-food applications (Scenario 4). Consumer acceptance of this technology will be key to determining which scenario becomes the future for each nation. The likelihood of each scenario is different for each nation, the U.S. will most likely evolve into scenario 2 or 3, while in the EU scenarios 1 or 4 are more likely. Determining the future for emerging nations is extremely complex and dependent on several risk factors, e.g., malnutrition rates and historical trading routes.

The future of biotechnology is complex. This analysis shows only four of the infinite options each nation can take in terms of accepting or rejection biotechnology for food consumption. The scenarios suggest that emerging nations are more sensitive to the timing of making this decision since most of these economies are directly connected to the US, EU or Japan, all of which may choose a different future for biotechnology.

Lastly, the impact this technology may have on small producers and rural areas needs further study. This analysis suggests that major changes will occur if this technology is virtually accepted. Small producers are ill prepared for a future where biotechnology is fully accepted. Additional data is required to capture the impacts this technology will have on small producers, rural areas and poor nations.

References

Akerlof, G., 1970. “The Market for “Lemons”: Quality Uncertainty and the Market Mechanism” *Quarterly Journal of Economics* 84(3), 488-500.

Brink, Johan. 2001. "Agricultural Biotechnology in South Africa: An Agricultural Research Council Perspective." Presentation at the Agricultural Biotechnology Support Project Seminar, East Lansing, Michigan, March 13.

Guichard, C., Delegate General of COLEACP, personal interview, May 2000.

Hill, Lowell and Battle. 2000. *Search for Solutions in the EU-US GMO Debate*. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, AE 4731.

H. C. Peterson and D. Weatherspoon, ”The Role of Scientific and Business Integrity in the Future of Biotechnology: A Scenario Analysis” Staff Paper 2001-17, Agricultural Economics Department, Michigan State University.

Schoemaker, 1995. "Scenario Planning: A Tool for Strategic Thinking”. *Sloan Management Review*, Massachusetts Institute of Technology, 36(2), 25-39.

Agriculture as a Key Issue for Rural Development in the European Union

Vidal, C.
Eurostat, Bâtiment Joseph Bech, 5 Rue Alphonse Weicker, L-2721 Luxemburg
e-mail: claude.vidal@cec.eu.int

Eiden, G., Hay, K.
CESD Communautaire, 3, Rue Wenceslas 1er, L- 2724 Luxembourg
e-mail: gerd.eiden@cesd.lu

Abstract: with a specific focus on agriculture, and using available data at Eurostat the main aim of this work has been to characterise the spatial components of rural areas within Europe. This data has been used to develop different typologies of 'rural', contributing to a fuller understanding of some of the similarities, differences and diversity between rural areas in regions and countries of the EU. It is anticipated that it will enable a comparative evaluation of territorial strengths, weaknesses, opportunities and threats. In this way, policy-makers and practitioners may use such information to formulate and target policies in a meaningful and regionally focused manner.

Keywords: agriculture, rural areas, rural development, indicators

1. Background – a change in direction for rural areas

Since the beginning of the 1980s structural policy in favour of rural areas has evolved substantially with numerous policy initiatives and changes being proposed and brought about by the European Commission. In particular, Community policy on rural development experienced an important change with the 1988 reform of the Structural Funds. Gradually policies have been taking on a greater territorial perspective with different measures such as environmental, agricultural and regional development increasingly horizontally integrated. The CAP reform after 1992, and the Treaty of Union confirmed that the development of rural areas must be included as part of Community policies for economic and social cohesion.

This shift in policy direction is due in part to the major restructuring occurring in rural areas (Barthelemy et al. 1999, Vidal 2000 a, b, c). These changes have helped contribute to the creation of a new image and role for areas, for example, in the conservation of the environment and protection of the natural heritage (European Commission, 1998). Agenda 2000 (European Commission, 1997) and the more recent Rural Development Regulation (European Commission, 1999) have continued to focus on the idea of horizontal integration of measures from a territorial perspective. Rural development became the second pillar of the CAP and takes into account environmental, marketing, tourism, rural services and village improvement. There was also an important shift to further 'green' what had formerly been regarded as agricultural structures' measures and agri-environment became a compulsory element in territorial rural development plans.

An in-depth analysis of the regional situation and specific conditions within rural areas is necessary as it is increasingly recognised that rural areas are diverse and undergoing significant changes. These need to be monitored, particularly for the purposes of future policy approaches.

It is impossible to use the term 'rural' generally as it holds very different meanings depending upon the geographical region. This diversity is reflected in a typology, which has been developed in the course of this study and based upon an extensive set of variables. It is anticipated that these typologies will be useful to policy makers and practitioners both when formulating new policies related to the development of rural areas and also as an indicator of how successful present policies might be within various regions.

2. Methodological approach

2.1 Rural areas: problems of delimitation

There have been several definitions of rural areas produced by different bodies over the years. One of the classification schemes from the late 1980s, and one that is also similar to more up-to-date classifications, was developed by the Council of Europe and incorporated aspects such as pressure of modern life, rural decline, very marginal areas and the level of integration within the national economy.

However, for administrative purposes rural areas are commonly categorised using single variable analysis. In 1994, the OECD (OECD, 1994) developed a simple territorial scheme that identifies three types of regions based on population density: predominantly rural, significantly rural and predominantly urbanised. Rural areas are defined as having population densities of below 150/km^2. The EUROSTAT (EC 1997) approach also distinguishes three different types of regions based on the degree of urbanisation: densely populated zones, intermediate zones and sparsely populated zones. Recent efforts in the framework of the 'Study Programme on European Spatial Planning' (SPESP, 2000) revealed that the inter-relationships and interdependence between urban centres and the rural hinterland are increasing, making any clear distinction even harder to formulate. The existence of so many different definitions highlights the difficulty when attempting to distinguish between rural and urban.

For the purposes of the report, however, it was decided to use population density alone as the distinguishing factor of 'rural' with a threshold of 200 inhabitants per km^2. Had the OECD definition of 150 inhabitants been used for example, over fifty percent of the NUTS 3 regions would have been excluded from the analysis. In some respects, therefore, 200 inhabitants per km^2 represented a compromise between the various definitions of 'rural' and the data available. From 455 NUTS regions at NUTS 2 or 3 level composing the entire territory of the EU 355 were included in this analysis.

2.2 Data sources and the geographic breakdown

In principal all variables available at Eurostat refer to the Nomenclature of Territorial Units for Statistics (NUTS [1995 and 1999]). In order to ensure that regions of comparable size are analysed, the statistical data refers to NUTS 3 level, except for Germany, Belgium and The Netherlands where data is related to NUTS 2 regions.

The report also had to take into account the fact that in January 1999 the Statistical Office of the EU introduced the latest modification to NUTS – NUTS 1999 which covers EU-15. It replaces the former NUTS 95, which only covered EU-12. The modifications to the territorial breakdown can have an impact on the consistency of the databases. Problems related to the modification can be summarised as follows:

- New coding: introduction of a new coding of the regions at different levels
- Disagregation: splitting of NUTS regions into smaller units, not existing before
- New territorial delimitation: due to a reorganisation of local governments etc. new NUTS regions are created.

Certain databases were updated, taking into account these modifications, others not. For the purposes of the report it was, therefore, necessary to find a certain territorial correspondence between the old and new regions. A table of convergence had to be created, linking the different databases and different territorial units.

2.3 Choice of variables for characterising rural areas

Based on the information available from Eurostat, and taking into consideration some of the problems with the data, a list of seven fields and associated variables were developed. These were chosen to reflect the development of rural areas but with a specific focus on the agricultural sector.

Demographic variables can be considered as some of the most important indicators when it comes to measuring strengths and weaknesses of a territory. Perhaps of greater interest, however, are the trends in population over time, and their causes and consequences. Changes in total population can be used as a measurement of development where population decline is an indicator of the economic fragility of an area or, on the other hand, where population increase is an indication of economic strength. This indicator can be reinforced further by taking into account age structure. The use of Gross Domestic Product (GDP) per capita (expressed in PPS [Purchasing Power Standards]) and evolution of GDP per capita was chosen as a measure of the regional value-added and welfare changes arising from economic activities in the region. It is seen as the best available estimate of average income levels in different regions and over time. It can also be used as an indicator of relative economic performance. Levels and changes in agricultural employment are important as they indicate the relative importance of agriculture in the local or regional employment structure, and how this is changing over time. As recent surveys prove, the level of agricultural employment is continually shrinking relative to output (due to both technological changes as well as structural changes). Therefore, through measuring the level of agricultural employment in the local economy, this will reveal how much employment needs to come from other sectors if the rural economy and population are to be sustained. The importance of agricultural employment is further developed through the inclusion of numerous variables associated with the farm labour force. Variables chosen within this field reflect the number of people employed (including regular family labour) per holding and per 100 ha, which can also be used as an indication of farm size, for example, whether or not a farm requires intensive labour such as those with permanent crops. It is also important to differentiate between labour types – whether or not farms have a high proportion of family labour force, part-time work, and seasonal work. Through the inclusion of both the age structures of those working within agriculture and the importance of off-farm work to their income, an insight is provided into the vitality and sustainability of the agricultural sector in an area.

Table 1: *Thematic fields and the list of corresponding variables*

Thematic fields	Variables
1. Demography	Population density
	Share of population less than 20 years
	Share of population older than 60 years
	Evolution of population
2. Economic Strength	GDP per capita and evolution of GDP per capita
	Unemployment rate and its evolution
3. Agricultural Employment	Share of agricultural employment and its evolution
4. Farm Labour Force	Average Number of Annual Working Units (AWU) by holding
	Average Number of AWU by 100ha
	Proportion of Family Labour Force (AWU) of total labour force
	Proportion of Young Holders (AWU) in the agricultural sector
	Proportion of Old Holders (AWU) in the agricultural sector
	Proportion of holders with other gainful activities
	Proportion of holders without spouses
	Proportion of old holders with other gainful activities
5. Agricultural Land Use	Proportion of Utilised Agricultural Area (UAA) of total land
	Proportion of Arable Land (from UAA)
	Proportion of Permanent Grassland and Pasture Land
	Proportion of Olive Trees
	Proportion of Vineyards
	Proportion of Permanent Cultures (except olive, vine)
	Proportion of Area Under Grass
	Proportion of Fallow Land
6. Farm Structure	Average Utilised Agricultural Area
	Average Economic Size of Holding
	Average Standard Gross Margin (SGM) per AWU
	Average Wooded Area of Holding
7. Livestock	Average Number of Bovines (LSU) per holding
	Average Number of Sheep (LSU) per holding
	Average Number of Pigs (LSU) per holding
	Average Number of Table Fowls and Laying Hens
	Average Number of Livestock Units per UAA
	Average Number of Herbivore per hectare Forage Crops, Permanent Meadows and Pastures
	Average Number of Carnivore (LSU) per hectare arable land and grassland

An analysis of the varying land uses in and between regions can be a valuable source of information for a number of reasons. It can be used as an indication of the different intensities of land use, how productive the land is as well as the diversity of land use. The fact that different crops are subject to different CAP commodity regimes under market policy, some knowledge of these can allow an assessment of changes in different regimes. Variables relating to livestock provide much the same information as those relating to land use. Both livestock indicators and agricultural land use, particularly those relating to intensity of production, are also useful in determining the environmental impacts of agriculture.

3. Statistical Analysis

In order to highlight the differences between rural regions of the European territory a statistical analysis of the variables had to be carried out. Figure 1 gives an overview of the different statistical procedures applied.

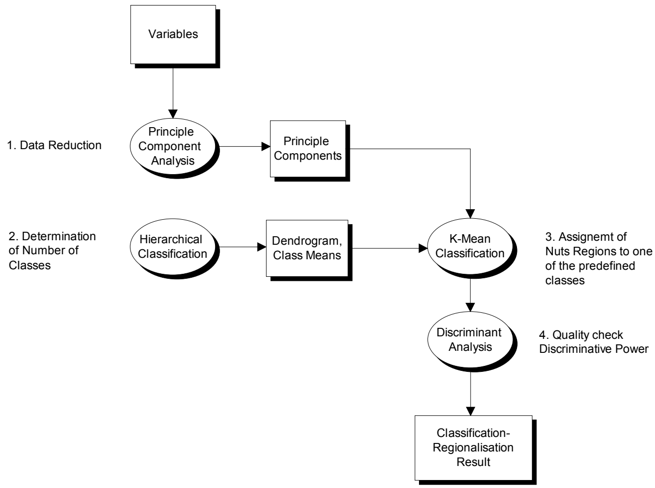


Figure 1: *Data processing chain*

(1) In the first stage a Principal Component Analysis (PCA) was carried out for each thematic field and the corresponding variables (see table 1). The aim of PCA is to define the underlying structure in a data matrix for the purpose of summarisation or data reduction. The principal components have been calculated for each of the thematic fields and the corresponding variables. In total 13 principal components were used, each of them clearly related to the original variables and thus enabling a meaningful interpretation of the components itself and their weight within the classification.
(2) The classification process started with a hierarchical cluster analysis aiming at a reasonable determination of the number of classes. By means of the dendrogram, a standard output of the agglomeration process, details about the similarity of the NUTS regions regarding the used PCA's are revealed, which eases the decision on how to group the different NUTS regions and facilitates a clear interpretation of the final classification result. A total of thirteen clusters - or types of rural areas - have been chosen. It is a compromise between the territorial detail required and to maintain characteristic differences between the clusters, which eases a clear and meaningful interpretation of the regional specificies.
(3) Based on the determination of the number of classes and the corresponding class means the final assignment of the NUTS regions to one of the predefined classes was done through a k-means cluster analysis.
(4) Finally, in order to evaluate the result of the classification procedure a discriminant analysis was performed. This procedure was used to validate the stability or reliability of the classification process. The quality check was aimed at detecting questionable assignments of certain NUTS regions to a cluster as well as the evaluation of the specific weight of variables within the final classification result.

4. Results and Interpretation: Typology of rural areas in the EU

As can be seen from the dendrogram, the cartographic presentation of final classification (map 1) result highlighted some remarkable features:
Particularly noticeable is the fact that the borders of many classes concur with those of the EU Member States. This is most obviously the case for the Nordic countries, Ireland and Portugal. The territorial pattern of the 13 groups shows a certain spatial consistency. When combined, the NUTS regions in nearly each of the 13 clusters are forming regional clusters (with the exception of class 5, 3 and 6). The 13 clusters can be combined without losing their consistency. However, whilst it was possible to aggregate the final thirteen classes into four general groups, there remains nonetheless, not only variation between the final four groups but also between regions within a group (perhaps with the exception of the Nordic Countries).

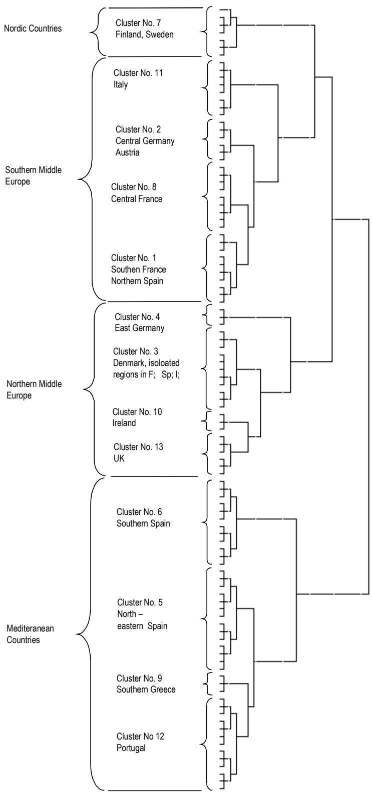
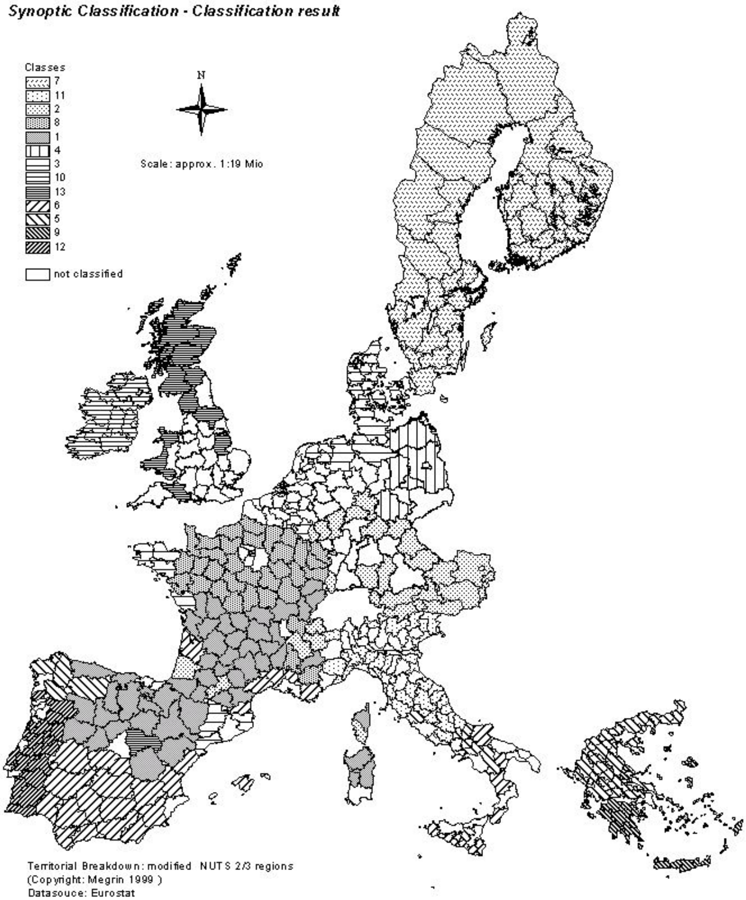


Figure 2: *Dendrogram of synoptic classification process (selection of regions)*

The following tables highlight the specificies and regional characteristics of the different types of rural areas in the EU. The Nordic regions were the only group which are made clearly of only one class. Southern Middle-Europe regions were made up of four classes which are quite similar in many respects, but at the same time there are important differences. All these regions have a below average labour force with the exception of Central Italy, below average economic size of farm, above average GDP (except S-France and N-Spain), a higher share of younger farmers (except Central Italy).
However, despite the similarities, it has to be noted that there are significant differences. Central Italy, for example has an ageing population with over a quarter of the population over 60 years of age. Unemployment rates also differ between these groups with Central France, Southern France and Northern Spain having much higher rates than the other classes within this group. As with the Southern Middle-Europe group, this group also had below average labour force but was characterised in particular by above average AWU per holding, high economic size of holdings (except Ireland) and low unemployment rates except in the East German Lander. Perhaps some of the more interesting data in this group relates to the East German Lander. This class is unique in many respects, not only in this group but also from the three other groups in the final classification.



Map 1: *Cartographic presentation of the final classification result*

This class had the highest unemployment rate, highest evolution of GDP, highest economic size of holdings, lowest rate of agricultural employment and the highest number of young holders compared to all other 12 classes, reflecting its position as a state in transition. The Mediterranean regions were possibly the easiest to associate together into one group and only had one or two minor differences as is shown in the table. These regions were characterised by small farm size, high levels of unemployment, an ageing population, low income levels and with agricultural employment dominating, particularly in Portugal where it reaches 20% of total employment.

Table 2: *Description of types of rural areas in EU*

Nordic region	Description	
Nordic Countries – Finland, Sweden (cluster No 7)	Below average farm labour force (1.4AWU/100ha)	High % of farmers with other gainful activities (52%)
	Low seasonal labour force (3.6%)	Low population density
	Above average % of young farmers (13.3%)	Population under 20 above average (25.3%)
	Above average economic size (18.8%)	Above average unemployment rate (13.3%)
	Above average farm size (28.2%)	
Southern Middle-Europe	Description	
Central Italy (cluster No 11)	Above average farm labour force (8.1AWU/100ha)	High % of young farmers with other gainful activities
	Lowest AWU per holding (0.8)	Highest population density (118.9)
	Below average % of young farmers (5.6)	Population over 60 above average (25.7%)
	Above average % of old farmers (60.6>55)	Below average unemployement rate (7.4 %)
	Below average economic size (9.5 ESU)	Below average farm size
	Below average unemployment (7.4%)	Highest level of GDP/capita
Central France (cluster No 8)	Below average farm labour force (3.4AWU/100ha)	Below average % of farmers with other gainful activities
	Above average AWU per holding (1.4)	Above average populaion density (81.2inh./skm)
	Above average % of young farmers (16.6<35)	Population under 20 above average (26.5%)
	Above average economic size (31.8 ESU)	Above average farm size
	Above average unemployment rate (11.6%)	Above average GDP
Austria, Lux, Germany, France, Itlay (cluster No 2)	Below average farm labour force (4.5 AWU/100ha)	Above average % of farmers with other gainful activities (43.8%)
	Average AWU per holding (1.0)	Above average population density (117)
	Below average economic size (13.9 ESU)	Below average farm size
	Above average % of young farmers (16.6 <35)	
	Below average unemployment rate (6.7%)	Above average GDP per capita
Southern France, Northern Spain (cluster No 1)	Below average farm labour force (3.5AWU/100ha)	Below average % of farmers with other gainful activities (25.7%)
	Average AWU per holding (1.1)	Below average population density (47.1)
	Below average economic size (13.3ESU)	Above average farm size
	Above average % of young farmers (15.1<35)	High % of permanent grassland/patures
	Above average unemployment rate (12.6%)	Below average GDP per capita
Northern Middle-Europe	Description	
Scotland and selected regions in UK (cluster No 13)	Below average farm labour force (2.2 AWU/100ha)	Above average % of old farmers with other gainful activities (20.6%)
	Above average AWU per holding (1.9)	Above average population density (94.7 inh./skm)
	Above average economic size (37.7 ESU)	Above average farm size
	Below average % of young farmers (8.4%<35)	High % of sheep (46.3), bovines (62.8) permanent grassland/pastures (56.2)
	Above average GDP/capita	Below average unemployment rate (7.4%) with decreasing tendency (-4.1%)
Ireland (cluster No 10)	Below average farm labour force (5.0 AWU/100ha)	Below average % of farmers with other gainful activities (25.1%)
	Above average AWU per holding (1.4)	Below average population density (48.9 inh./skm)
	Below average economic size (12.9 ESU)	Above average farm size
	Above average % of young farmers (13.3%)	Highest % of permanent grassland/pastures (80.1)
	Below average GDP per capita	High % of LSU per AWU (1.5%)
	Low unemployment rate (9.9%) with decreasing tendency (-6.8)	
Denmark, Northern parts of Germany (cluster No 3)	Below average farm labour force (4.4 AWU/100ha)	Below average % of farmers with other gainful activities (25.5%)
	Above average AWU per holding (1.3)	Above average population density (105.9)
	Above average economic size (34.8 ESU)	Above average farm size
	High % of LSU per UAA (1.9)	High % of arable farming (65%)
	Low % of agricultural employment	Low unemployment rate (7.3%)with decreasing tendency
East German Lander (cluster No 4)	Lowest farm labour force (2.0 AWU/100ha)	Above average % of young holders with other gainful activities (46.4%)
	Highest AWU per holding (4.3)	Above average population density
	Highest average economic size (112.3 ESU)	Lowest rate of agricultural employment
	Above average number of young holders (21.7%)	High % of bovines and sheep
	Below average GDP/capita	Highest unemployment rate (19.2%)
	Highest evolution of GDP/capita (14%)	

Mediterranean Countries	Description	
Mediterranean Regions (cluster No 6,5,9,12)	High AWU/100ha	High share of part-time labour
	Below average AWU per holding	Small % of young farmers
	High proportion of older farmers	Low GDP per capita
	High unemployment rate	High proportion of agric. employment
	Domination of permanent crops	
Southern Spain, South coast France **(cluster No 6 only)**	Low AWU/100ha	High unemployment rate (22%)
	High agricultural employment	High proportion of seasonal labour force (26%)
NW Spain, Greece, Italy **(cluster No 5 only)**	Highest unemployment rate (12%)	
Portugal **(cluster No 12only)**	Highest share of agricultural employment 20%	High evolution of unemployment rate 6.2%
	Small average farm size (11ha)	

5. Conclusions

The classification obtained illustrates the great diversity of rural areas in the EU and showed some remarkable spatial differences between regions. Although a great deal of variation exists between areas, there is nonetheless a distinct regional and spatial coherence, which can be explained in a number of ways. The coherence indicates that rural areas follow geographic and national specific characteristics. This is proved when aggregating the thirteen clusters to four overall groups, resulting in a clear north-south gradient. This would seem to follow from the fact that the variables chosen are a result of both historical and social development (farm structure, economic strength etc) and physical geographical conditions (e.g. agricultural land use). A second element of the explanation may be related to the quality of the data used. Whilst every effort was made to harmonise the data, definitions for certain variables vary considerably between member states of the EU and these national specific definitions could have had some influence upon the final classification result. With regard to the thematic interpretation of the classification result, it is important to be aware of the fact that certain variables played more of a role in the final regionalisation results. These were farm labour force, farm structure, and the economic strength of a region. Despite the fact that only a limited number of the 39 original variables were included in the final analysis (i.e. the principal components), it was, nonetheless, clear that those relating to the agricultural sector (farm structure, agricultural employment and farm labour force) contributed substantially to the diversity of rural regions. However, with regard to the choice of variables, some improvements are necessary, particular considering other relevant and important characteristics of rural areas such as remoteness, accessibility or social integrity etc.

These findings are of particular relevance to the continuing debate and policy development concerning rural areas. They highlight the fact that, although rural areas within a region may contain certain similarities, there are significant differences between regions themselves. As a result, policy must continue to be territorially specific rather than sector specific. For example, the importance of agriculture to the economy varies enormously between regions. In terms of employment, it is far more important to the Mediterranean regions than those in the North. Such diversity between regions, therefore, cannot be ignored. If the issues concerning rural areas are to be addressed in an appropriate and coherent manner then typologies such as the one conducted for the purposes of this report are extremely important. This territorial classification makes an important contribution to the identification of the strengths, weaknesses and opportunities in rural areas across the European territory and it is hoped that it will be used by the Commission as complementary or background information.

References

Barthelemy, P.A.; Vidal, C. (1999) *Rural realities in the European Union*, in European Commission (1999) *Agriculture, Environment, Rural Development - Facts and Figures: A Challenge for Agriculture*, Brussels

Council Regulation (EC) No 1257/1999 of 17 May 1999 on Support for Rural Development. EC.

European Commission (1997) *Rural Developments*, CAP 2000 Working Document

European Commission (1998) *The Agricultural Situation in the European Union*, 1998 Report, Brussels

European Commission (1999) *Methods and Nomenclatures - NUTS*, Luxembourg: Office for Official Publications of the European Communities

European Commission (2000) *Agenda 2000: Development of the Common Agricultural Policy* (CAP), Brussels.

European Commission (2000): *Study Programme on European Spatial Planning*. Final Report at www.nordregio.se

OECD (1994) *Creating Rural Indicators for Shaping Territorial Policy*, OECD, Paris

Vidal, C. (2000a) *Thirty Years of Agriculture in Europe: Farm Numbers Declining as Farms Grow in Size*, Eurostat-Statistics in Focus, Agriculture and Fisheries, Theme 5 – 3/2000

Vidal, C. (2000b) *Ever Larger Holdings but different Economic Situations*. Eurostat-Statistics in Focus, Agriculture and Fisheries, Theme 5 – 9/2000

Vidal, C. (2000c) *Arable Land, increased Size of Holdings and Changes in Crop Rotation.* Eurostat-Statistics in Focus, Agriculture and Fisheries, Theme 5 – 12/2000

5 June 2001

Invited Paper Session

THE AGRICULTURE-ENVIRONMENT RELATIONSHIP: TREND AND FEEDBACK

Chairs: A. Stein - P. Monestiez

The "Precision" Approach to the Agricultural Production-Environment Relationship ([1])

Alex B. McBratney
Australian Centre for Precision Agriculture,
McMillan Building A05, The University of Sydney, NSW 2006, Australia
e-mail: Alex.McBratney@acss.usyd.edu.au

Abstract: Precision agriculture (PA) may be defined as an integrated information- and production-based farming system that is designed to increase long-term, site-specific and whole-farm production productivity and profitability while decreasing negative impacts on the environment. Its principal guise is site-specific crop management (SSCM). Achieving a viable PA management system requires the gathering of detailed information within fields. Here I focus on a few statistical aspects. In section 2, I present a, largely statistical, way of measuring the opportunity for SSCM. An important key layer of information for economically and environmentally optimising is knowledge of whole-field and within-field fertiliser response functions. Field experiments to obtain such response functions are in their infancy. These are discussed very briefly in section 3. In section 4, I wrestle with the difficult problem of quantifying environmental impacts economically and spatially, which is required for proper systems optimization.

Keywords: precision agriculture, environmental economics, field experimentation

1. Introduction

Precision agriculture (PA), a relatively recent concept, may be defined as, 'an integrated information- and production-based farming system that is designed to increase (or maintain) long-term, site-specific and whole-farm production productivity and profitability while minimising (or decreasing) negative impacts on the environment.' Its principal guise is site-specific crop management (SSCM). Achieving a viable PA management system requires the gathering of detailed information within fields. In this talk I focus on some statistical aspects. In section 2, I present a, largely statistical, way on measuring the 'opportunity' for site-specific crop management.

A key trade-off is the replacement of physical, chemical and biological inputs by information – the latter having no environmental impact. Various 'high' technologies have been developed for gathering this information, and a number of quantitative approaches such as spatio-temporal and mechanistic models is required to process the information for decision-making. An important key layer of information for economically and environmentally optimising inputs in site-specific crop management is

([1]) The present paper is financially supported by the Australian Cotton Research & Development Corporation, the Australian Cotton Cooperative Research Centre, the Australian Grains Research & Development Corporation and the University of Sydney. I should like to thank my colleagues Matt Pringle, Craig Stewart and Brett Whelan for the provision of valuable data.

knowledge of whole-field and within-field fertiliser response functions. Field experimentation is required to achieve this – but classical approaches may be inefficient in time and money. Field experiments to obtain site-specific response functions are in their infancy. These are discussed very briefly in section 3.
In section 4, I try to grapple with the enormously difficult problem of quantifying environmental impacts economically. This is required for proper system optimisation.

2. Opportunity

Whether one decides to embark upon site-specific crop management depends on whether there is an economic or management opportunity. Such opportunity can be gauged by looking at the spatial patterns of yield components. Very high resolution information on yield is now available from electronic yield monitors in harvesting for a number of crops. Yield estimates from remote-sensing imagery can also be used. According to McBratney et al. (2000) and Pringle et al. (submitted) the opportunity for SSCM can be regarded as a function of a Magnitude component, an Area component, and Economic/environmental concerns. The SSCM Opportunity Index (OI) can be defined as:

$$\mathrm{OI} = \sqrt{\left(\frac{\text{auto - ARCOVAR}_{1000}}{Median(\text{auto - ARCOVAR}_{1000})}\right) \times \left(\frac{pT \times A + (1 - pT) \times J_a}{MZEM}\right) \times E} \; . \qquad (1)$$

Magnitude ↑ (first term); Area ↑ (second term); Economics / environment ↑ (E)

2.1 Magnitude

The relative magnitude of yield variation could be found by comparing CVs to a median value, however, we disagree with the use of a standard CV in this situation. Firstly, the CV is non-spatial and therefore potentially misleading when dealing with different sized areas. Secondly, the CV tells nothing of the difference between autocorrelated yield variation (which is manageable), and inherently random ('nugget') variation (which is not manageable). The CV is therefore undesirable, and a better method of describing magnitude is needed.
'ARCOVAR' is an acronym for 'AReal COefficient of VARiation', which is a method of standardising the previously non-spatial CV to an area. ARCOVAR is based on the double integral of the yield variogram, although the C_0 parameter was excluded from the integration because we are only interested in autocorrelated variation (hence the prefix 'auto'). The ARCOVAR procedure is outlined here.
Variograms are made of the raw yield of each field and fitted, weighted by m at each lag with the exponential, spherical, double exponential and double spherical theoretical variogram models fitted. These theoretical models have sills (i.e., a finite ceiling on their maximum variance). If there is a trend in the variogram (i.e., no obvious sill), the maximum possible value of R_2 is constrained to a maximum of 1000 m, which thereby forced a sill upon the variogram. Model performance is assessed by the Akaike Information Criterion (AIC) Webster & McBratney, 1989):

$$\text{AIC} = n \times \ln\left\{\sum_{i=1}^{n}(\text{empirical} - \text{theoretical})^2\right\} + 2 \times p, \tag{2}$$

where, n is the number of lag distances on the empirical variogram, and p is the number of parameters in the theoretical model. The lower the AIC, the better the model's performance. The AIC penalises a model's goodness-of-fit against its number of parameters. The exponential and spherical models have three parameters, while 'double' models have five. Parameters of the theoretical variogram model with the lowest AIC are (numerically) double-integrated (minus the C_0 parameter) to the standardising area (chosen as 1000 ha). The double integration of variogram parameters is given by (after Webster & Oliver, 1990):

$$\bar{\gamma}(1000\ \text{ha}) = \frac{1}{B^2}\int_B\int_B\{\gamma(\mathbf{x},\mathbf{x}') - C_0\}\,d\mathbf{x}\,d\mathbf{x}', \tag{3}$$

where:

$\bar{\gamma}(1000\ \text{ha})$	=	average variance of yield within 1000 ha;
B	=	the number of points used to discretise the 1000 ha area;
$\gamma(\mathbf{x},\mathbf{x}')-C_0$	=	the semivariance between any two of the points that discretise the block, minus the nugget variance.

The square root of $\bar{\gamma}(1000\ \text{ha})$ was then divided by the field's mean yield and multiplied by 100 to obtain the auto-ARCOVAR$_{1000}$:

$$\text{auto-ARCOVAR}_{1000} = \frac{\sqrt{\bar{\gamma}(1000\ \text{ha})}}{\mu(\text{yield})} \times 100 \tag{4}$$

The median auto-ARCOVAR$_{1000}$ of all the fields was found, and used as the quantity against which to compare the magnitude of autocorrelated yield variation. Therefore, in (1), the division of auto-ARCOVAR$_{1000}$ by its median effectively states that the opportunity for SSCM will be increased if a crop is more variable than what is usually observed.

2.2 Area

In the Introduction, it was implied that trend in a yield map is desirable when we wish to implement SSCM. However, the manifestation of trend in a yield map's variogram will imply that the average autocorrelation area of yield is infinite, which will lead to extremely large (and unrealistic) OI values. (Trend was not considered a problem when modelling variograms for the Magnitude component of (1), because a finite area – 1000 ha – was being used to find a value of yield variation; here, this process is essentially being reversed).

To calculate the average area within which yield is autocorrelated, yield data must be decomposed into a quartic trend-surface and an error term:

$$Y(E,N) = \begin{pmatrix} \text{Int.} + E + N + E^2 + N^2 + E\times N + E^3 + N^3 + E^2\times N + \\ E\times N^2 + E^4 + N^4 + E^3\times N + E^2\times N^2 + E\times N^3 \end{pmatrix} + \varepsilon \tag{5}$$

where:

E	=	Easting coordinates of yield;
N	=	Northing coordinates of yield;
Y(E,N)	=	yield as a function of its Eastings and Northings;
Int.	=	intercept of regression;
ε	=	error term (residuals).

Although it depends on the size of the field, yield monitor data is usually so abundant that a reasonably complex trend model can be afforded, hence a fourth-order model has been used here.

Empirical variograms were then made of the residuals, and fitted with the four theoretical models given in the Appendix. The best-fitting model was again found by the AIC (Equation 5). This model of spatial variation was then used to find the 'areal scale' of the yield residuals (J_a). Russo & Bresler (1981) employed this concept to determine the spatial dependence of soil hydraulic properties. We have adapted their idea to approximate the average area within which the residuals of a yield trend-surface are autocorrelated (after Russo & Bresler, 1981):

$$J_a \approx \frac{\left\{2\int_0^\infty \left(1-\frac{\gamma(h)}{(C_0+C_1+C_2)}\right)hdh\right\}}{10000}, \tag{6}$$

where:

$\gamma(h)$	=	theoretical variogram of yield residuals as a function of h;
C_0, C_1, & C_2	=	parameters of the residual's theoretical variogram (if the best fitting model was not a 'double', C_2 was equal to nought);

Equation 9 specifies that the best-fitting residual variogram model be converted into an equivalent correlogram. This procedure requires that the variograms have a sill. The presence of trend in a variogram will often make it increase infinitely (i.e., as if it has no sill), such that a sill must be forced upon the variogram. Hence the reason why the residuals from a quartic trend surface have been used to find J_a, and also why only theoretical models with finite sills (i.e., exponential and spherical models) have been fitted.

The quantity pT represents the proportion of total yield variance explained by the quartic trend-surface. Because a trend-surface is theoretically autocorrelated to an infinite area, a limit must be employed; this was chosen as the area of the field (A). Multiplying pT by A gives the contribution of the trend surface to the average area within which yield is autocorrelated, while multiplying J_a by (1-pT) gives the contribution of the residuals.

The Minimum Zone of Effective Management (MZEM) is an estimate of the minimum area (in hectares) within which variable-rate controllers can reliably operate. It is calculated as:

$$MZEM = \frac{(\beta\nu\tau)}{10000}, \tag{7}$$

where:

β = width of application swath (m)

ν = speed of vehicle (m/s)
τ = time required to alter application rate (s)

So, the division of $pT \times A + (1 - pT) \times J_a$ by MZEM in (1), effectively states that the opportunity for SSCM will be increased when farm machinery can operate within the average area within which yield is autocorrelated; if this is not the case then SSCM is hardly feasible.
McBratney et al. (2000) and Pringle et al. (submitted) present a variety of results on the calculation of the Opportunity Index.

3. High technologies – yield mapping and variable-rate experimentation

Precision agriculture has to a degree been driven by the advent of various high technologies, particularly the coupling of real-time positioning using GPS, electronic yield monitors on harvesting machines, proximal soil sensors, variable-rate controllers for fertiliser, pesticide and seed applicators and GIS (geographic information systems). The coupled technology provides the data to investigate opportunity and to manage inputs differentially. Yield maps such as the one shown in Fig. 1, which because of their high data intensity have their own statistical issues, and the equations in section 2 are a key basis for the process. Monitors are available for small grains, cotton, sugar cane, grapes and a variety of horticultural crops.

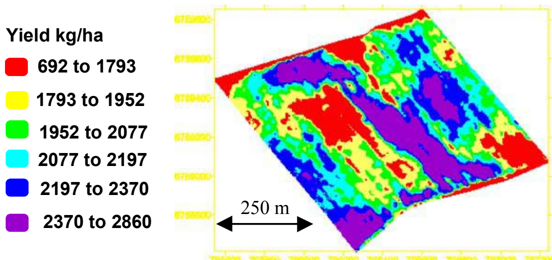


Figure 1: *A cotton-lint yield map for a field in northern New South Wales. The map is produced from some 40000 observations from a yield sensor and then interpolated onto a grid using ordinary kriging with a local variogram.*

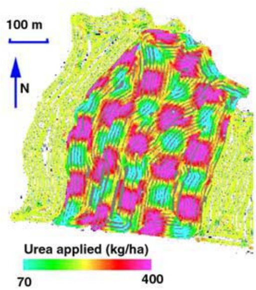


Figure 2: *Spatial distribution of ureas applied using a variable-rate controller in a field in northern new South Wales. The pattern approximates an egg-box design.*

One of the key issues is to develop strategies for within-field management. This can be either through management zones or continuous moving-window management. Take the

case for, e.g., fertilisers, here we would like to know the response function for different management zones within the field or as a continuous moving window function of spatial coordinates. The variable-rate technology allows the setting down of sophisticated field experiments in farmers' fields to acquire this knowledge (Fig. 2) and the yield-monitoring technology allows the measurement of response. The spatial design of such 'on-farm' experiments are in their infancy. Adams & Cook (2000) discuss some possibilities. For management zones, an efficient design would seem to be the "*fleck*" design where randomised block experimentation is done with spatial constraints and economic considerations. An economic consideration is that one does want to penalise the grower by using sub-optimal rates over much of the field. Most of the field can have a uniform treatment which the grower considers normal. Data from all of the field can be used in the analysis. The proper objective function and design for these experiments have yet to be developed, but an approach homologous with the use of spatial simulated annealing (Van Groenigen & Stein, 1998) for spatial sampling, seems the most obvious one.
Where the object is to produce a local moving window response function, and one is not concerned about the grower's production from the experimental field, systematic designs such as a "*draught board*" with four levels or an "*egg-box*" design as shown in Fig. 2, sometimes called a two-dimensional sine wave (e.g., by Adams & Cook, 2000) can be used.
The power of the technology, coupled with the goal of improved management, provides many new possibilities and challenges for field experimental design and analysis. Ultimately, this will provide increased quantity of information about environmental responses within fields.

4. Economic/environmental benefit

At present, little is known about the nature of parameter *E* in (1). Some of the factors that *E* must consider will be short- and long-term economic goals, and the on-farm and off-farm impact of management practices. Economic considerations in precision agriculture have thus far been concerned with the cost of implemenation versus increased value of production (e.g., Bongiovanni & Lowenberg-DeBoer, 2000; Lowenberg-DeBoer, 2000), or more importantly, the value of (increased) information (Bullock & Bullock, 2000), but has not dealt with the joint production-environmental question however. Previous presentations of the Opportunity Index (McBratney et al., 2000; Pringle et al., submitted) have evaded the evaluation of *E*. Here I present a tentative first step towards coupling environmental costs and benefits with production costs and benefits as a formal site-specific optimisation target, the true goal of precision agriculture. The kind of thinking and analysis below is becoming known under terms such as 'valuing ecosystem services' or 'environmental capitalism'.
A conventional model, or production function for nitrogen fertilization is of the form,

$$Y_r = \alpha + \beta.N + \gamma.N^2, \qquad (8)$$

where Y_r is the relative yield in Mg/ha, N is the amount of fertiliser N applied in kg/ha, and α, β and γ are parameters. For example, $\alpha = 0.17$, $\beta = 0.01079$ and $\gamma = -0.0000341$. The actual yield Y (Mg/ha) is given by,

$$Y = Y_c.Y_r = Y_c(\alpha + \beta.N + \gamma.N^2), \qquad (9)$$

where Y_c is the yield coefficient. For wheat somewhere in SE Australia, it would not be unreasonable for Y_c=5. The conventional economic model takes into account the value of production and the costs of applied fertiliser, i.e.,

$$G\$ = V\$ - C\$ \qquad (10)$$

where *G$* is the gross margin in $/ha, *V$* is the value of production ($/ha) and *C$* is the cost of fertiliser ($/ha). The equation can be expanded to give,

$$G\$ = \{Y_c.P\$(\alpha + \beta.N + \gamma.N^2)\} - \{N.F\$), \qquad (11)$$

where *P$* is the unit value of production ($/Mg) and *F$* is the unit cost of fertiliser ($/kg). In Australia, *P$* = A$165 and *F$* = A$1.15. The normal economic approach is to find the N rate, N_{opt} that maximises *G$*. This is found by differentiating and rearranging (11) to give

$$N_{opt} = -((\beta Y_c.P\$) - F\$)/(2.\gamma\, Y_c.P\$) \qquad (12)$$

This gives an N_{opt} = 134.6 kg N/ha, which maximises *G$* at A$662/ha. For this amount of applied N the production Y is 4.95 Mg/ha. If one uses a pure production model, i.e., assume (or pretend) that the cost of fertiliser *F$* in (12) is zero, then the optimal N is the maximum yield 5 Mg/ha, which occurs for N_{opt} = 154.6 kg N/ha, and a G$ of A$650/ha. (These are presented in Table 1 below.)

4.1 Environmental economic model

We can now extend the usual fertiliser application economics to deal with the environmental costs of over-fertilising. The basic idea is that the cost of recovering fertilizer N not used by the crop and lost to the environment by leaching or denitrification has to be taken into account. It has been known for at least 40 years (since the first use of ^{15}N mass spectrometry) that it is usual for only 50% or less of applied N to appear in the crop – the so-called ‘enigma of nitrogen balance sheets’. Let’s call the proportion of applied N lost to the environment the nitrogen inefficiency coefficient I, we shall call its complement the nitrogen efficiency coefficient, i.e., I =1-E. What is the environmental cost of the nitrogen ‘lost’ to the environment? One approach is to calculate the cost of removing it from the environment. We can estimate this as a coefficient of the cost of fertiliser. Professor Louis-Marie Bresson, Departement Agronomie-Environnement, Institut National Agronomique Paris-Grignon (personal communication) estimates in France the clean-up coefficient C_c is 4 (times the cost of fertiliser). We’ll assume an average but variable clean-up coefficient for loss as leached nitrate and atmospherically as N_2O and NO_2. These environmental effects can be accommodated in (10) by adding another term, i.e.,

$$G\$ = V\$ - C\$ - E\$, \qquad (13)$$

where *E$* is environmental clean-up cost of N lost to the environment in $/ha. This is given by,

$$E\$ = C_c.I.F\$.N_{opt} \quad (14)$$

Of course (13) can be written as *G$* = *V$* - (*C$*+*E$*) = *V$* - *T$*, where *T$* is 'total' cost of fertiliser. Similarly to (11), (13) can be differentiated and rearranged to give

$$N_{opt} = -((\beta Y_c.P\$) - ((C_c.I)+1).F\$)/(2.\gamma\, Y_c.P\$) \quad (15)$$

The diagram (Fig. 3) below shows N_{opt} for a range of C_c and E.

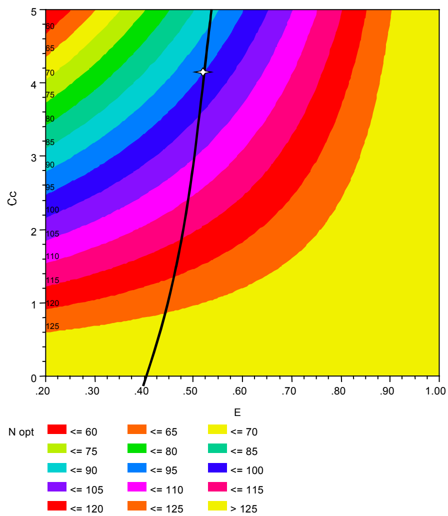


Figure 3: *Contours of optimal N application in kg/ha (N_{opt}) as a function of nitrogen-use efficiency, E, and clean-up coefficient, C_c. The solution locus for (16) substituted into (15) is shown as the black (almost vertical) line. The star is located at (E=0.5, (I=0.5), C_c=4), see also results in Table 1.*

Table 1 reports the results for N_{opt}, the yield and the gross margin ($/ha) for a pure production model, a model that includes the cost of fertiliser, and an environmental model with a reasonable choice of C_c and E (the point shown as a star in Fig. 3). Clearly, for this example, introducing an environmental clean-up penalty significantly affects the amount of fertiliser used, production (with its food security implication) and gross margins. The third scenario (the environmental economic one) assumes the user of the fertiliser pays for the clean-up, therefore the nett environmental cost to the community is zero, as opposed to environmental penalties to the community of $356 and $310 per hectare for the first two scenarios.

Table 1: *Optimal N application rates, yields, gross margins and environmental costs for three economic scenarios.*

Optimality criterion	N opt	Yield at N opt	G $ at N opt	E$[#]
Max. Production (Y_c=5)	154.6	5.00	650	356
Economic, including cost of fertiliser	134.6	4.95	662	310
Economic, including cost of fertiliser + cost of environmental clean-up (I=0.5 , C_c=4)	94.7	4.39	398	218

4.2 Extensions

The model is clearly somewhat simplistic as one would suspect a negative interaction between efficiency and amount of fertilizer applied, e.g., for optimal N,

$$E = 1 - 0.007228.N_{opt} + 0.00002030.N_{opt}^2 \quad (16)$$

When (9) is substituted into (8), the solution locus is shown as the black line in Fig. 2. (Many other improvements could be made to the nitrogen-balance model but that is beyond the scope of this study. It is the concept that is important.)

4.3 Implications and applications

What are implications of this approach for precision agriculture in particular and agriculture in general? Accepting the environmental economic framework, precision agriculture is required to:-

(i) Lessen the negative impacts of its application on production and profitability The situation is probably improved if N application is optimised locally. (See Fig. 4)

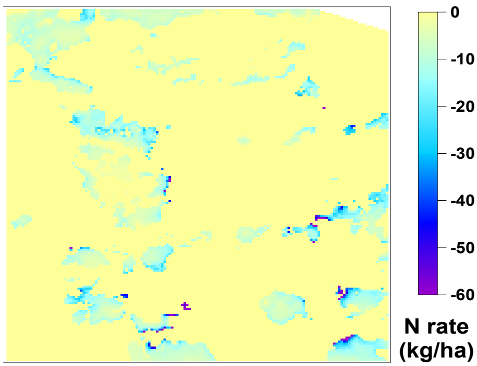


Figure 4: *Spatial distribution of environmentally-economic optimal N rate (15) minus economic N rate (12) for a 800 m by 800 m wheat field in Western Australia assuming I=0.5 and Cc=4 etc. The results were obtained by fitting local moving-window quadratic functions to N applied according to a draught-board design*

(ii) Maximise E and optimise applied N site-specifically. Within a field Y_c and E will vary spatially. The comparative advantage of the precision approach will depend on the amount and structure of that spatial variation. (See Fig. 4)

[#] For the first two scenarios E$ is estimated assuming I=0.5,Cc=4, E$ = 2.3Nopt

(iii) *Environmentally audit.* Yield and protein mapping along with maps of fertiliser application maps will allow growers to environmentally audit their production which is a necessary part of the application of the environmental economic model for Precision Agriculture.

5. Conclusion

Clearly, the reasonable and inevitable environmental economic model, given in Eqs. (13), (14) & (15), optimised at every point in the field, provides strong challenges to, and a powerful rationale for, precision agriculture. The data to solve such equations are potentially available through the technology of precision agriculture. Much has still to be done on devising new intensive spatio-temporal statistical models for design, sampling and analysis to transform the data into meaningful information.

References

Adams, M.L., Cook, S.E. (2000) On-farm experimentation: application of different analytical techniques for interpretation, Paper 4, in: *Proceedings of the 5th International Conference on Precision Agriculture,* Robert, P.C., Rust, R.H. & Larson, W.E. (Eds.), SSSA/CSSA/ASA, Madison, Wisconsin. 17 p.

Bongiovanni, R., Lowenberg-DeBoer, J., (2000) Economics of variable rate lime in Indiana. *Precision Agriculture*, 2, 55-70.

Bullock, D.S., Bullock, D.G., (2000) From agronomic research to farm management guidelines: a primer on the economics of information and precision technology. *Prec. Agr.*, 2, 71-101.

Lowenberg-DeBoer, J., (Ed.) (2000) *Precision Farming Profitability.* Agricultural Research Programs, Purdue University, West Lafayette, Indiana, USA. 132 pp.

McBratney, A.B., Whelan, B.M. , Taylor, J.A. , Pringle, M.J. (2000) A management opportunity index for precision agriculture, Paper 207, in: *Proceedings of the 5th International Conference on Precision Agriculture,* Robert, P.C., Rust, R.H. & Larson, W.E. (Eds.) SSSA/CSSA/ASA, Madison, Wisconsin.16 p.

Pringle M.J., McBratney, A.B., Whelan, B.M., Taylor, J.A. (submitted) A preliminary approach to assessing the opportunity for site-specific crop management in a field, using yield monitor data. *Agricultural Systems*.

Russo, D., Bresler, E. (1981) Effect of field variability in soil hydraulic properties on solutions of unsaturated water and salt flows. *Soil Science Society of America Journal*, 45, 675–681.

Stewart, C.M., McBratney, A.B. (2000) Development of a methodology for the variable-rate application of fertiliser in irrigated cotton fields, Paper 180, in: *Proceedings of the 5th International Conference on Precision Agriculture,* Robert, P.C., Rust, R.H. & Larson, W.E. (Eds.), SSSA/CSSA/ASA, Madison, Wisconsin. 13 p.

Van Groenigen, J.W., Stein, A., (1998) Spatial simulated annealing for constrained optimisation of spatial sampling schemes. *J. Env. Quality*, 27, 1078-1086.

Webster, R., McBratney, A.B. (1989) On the Akaike Information Criterion for choosing models for variograms of soil properties. *Journal of Soil Science*, 40, 493-496.

Webster, R.., Oliver, M.A. (1990) *Statistical Methods in Soil and Land Resource Survey.* Oxford University Press. New York, NY.

Statistical Issues for Water Quality at Different Scales in Agricultural Regions

D. J. Mulla, P. H. Gowda, A. S. Birr
Dept. Soil, Water, and Climate, University of Minnesota
1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108, USA
e-mail: dmulla@soils.umn.edu

Keywords: Water quality, scale issues, agriculture, nitrogen.

1. Introduction

Non-point source pollution from agriculture is a widespread problem in Europe and North America. Concerns typically include the nutrients nitrogen and phosphorus, as well as herbicides and pathogens. Researchers working on transport and fate of agricultural nutrients have typically focused on the spatial scale of the small plot, hillslope, or field, at the temporal scale of several months or years. The aim of most researchers is to describe the processes and pathways that control the concentration and load of nutrients leaving the plot, hillslope, or field in response to various factors that might include climate, soil, or landscape properties, and management alternatives.

In contrast, agency leaders and policy makers are typically concerned about decade or century long regional trends in water quality at the scale of watersheds, basins, counties, provinces, states, and countries. They are typically interested in monitoring regional trends in water quality, identifying the level of impairment, identifying the sources of pollution, and developing goals and strategies for restoring good water quality.

A few researchers have started to focus their research at the spatial and temporal scale of greatest concern to agency leaders and policy makers. They have encountered significant challenges arising from an inability to model and collect data at the fine resolution needed for evaluation of regional water quality trends. This paper provides an overview of some of these challenges, and gives examples being used to address nitrogen transport and fate at various scales in Minnesota.

2. Scales and Scale Issues

The scales at which transport and transformation of nutrients occur are diverse. Spatial scales of typical interest include the pore, pedon, plot, hillslope, field, minor watershed, county, major watershed, basin, state, or country. Temporal scales include the picosecond, microsecond, second, minute, hour, day, month, year, decade, century, millenium, or eon.

Typically, we are interested in spatial and temporal scales that are congruent. Studies of nitrogen transformation kinetics at the pore scale occur in milliseconds, rather than eons. Studies of nitrogen transport at the watershed scale occur in days or months. Changes in global soil organic nitrogen and carbon sequestration occur in decades or centuries. It is rare to study incongruent spatial and temporal scales, for instance processes at the pore scale over a time scale of centuries. Or studies of nitrogen loads to the Gulf of Mexico at the scale of milliseconds.

Scale issues have been discussed in several previous papers (Addiscott et al., 1995; Jury and Gruber, 1989; Kirby et al., 1996; Loague et al., 1996; Mulla and Addiscott, 1999; Wagenet and Hutson, 1996). Some of the more important issues involve spatial and temporal variability, complexity, non-uniqueness, non-linearity, patchiness, and change in process. Various approaches for up-scaling, down-scaling, combinations of up- and down-scaling, and strategic cycling have been developed (Root and Schneider, 1995).

The appropriateness of statistical methods change with scale. At the pore scale, statistical mechanics might be appropriate, but this is not true at the watershed scale. At the plot and field scale, controlled experiments involving variation of nitrogen rates and measurement of water quality impacts are typically evaluated using analysis of variance. The latter experiments would not be possible at either the pore or basin scales. At the watershed scale, experimental replication becomes difficult. Paired watershed studies may be used to study water quality impacts of varying nitrogen application rates at this scale. Either geostatistical or regression methods are appropriate at the hillslope scale and at coarser resolutions.

Temporal statistical methods traditionally involve time series analysis. Examples include smoothing techniques, forecasting techniques, and auto regressive techniques. Combined spatial and temporal statistical tools are rare. Examples include state-space analysis and wavelet analysis (Shumway et al., 1989; Lark and Webster, 1999).

3. Processes, Pathways, and Data Needs: Example of Nitrogen

Nitrogen is an essential plant nutrient. In agricultural regions, addition of nitrogen to growing crops often leads to significant increases in yield. Addition of excess nitrogen can lead to serious water quality degradation. Determining what constitutes an excess application rate is complicated because of the numerous processes and pathways that affect the fate and transport of nitrogen. These include plant uptake, nitrification, denitrification, mineralization, immobilization, runoff, erosion, and leaching or drainage. All of these are subject to both spatial and temporal variability.

In addition to the processes and pathways, it is essential to have accurate information concerning the inputs of nitrogen from fertilizer, manure, atmospheric deposition, and fixation. At the scale of plots, hillslopes, and fields, these inputs can be reasonably controlled through management. At the scale of watersheds and large regions, we must increasingly rely on statistical survey information for the sales of fertilizer, number and species of farm animals, average rates of manure production and manure nutrient content, and types of confinement, storage, or land application methods.

Data needs for evaluating nitrogen fate and transport must typically address these pathways. As the scale of study becomes coarser, it becomes more difficult to obtain accurate information concerning nitrogen fate and transport pathways and processes. For example, what is the spatial and temporal variability in denitrification at the scale of a major watershed? Does it matter if we use a spatially or temporally average denitrification value across the entire watershed?

At coarse scales, information concerning spatial and temporal variations in precipitation become important. Other useful information includes landuse, dates of crop planting and harvest, slope steepness, residue cover, and soil properties (drainage, permeability, organic matter, etc).

4. Nitrogen Losses at Various Scales

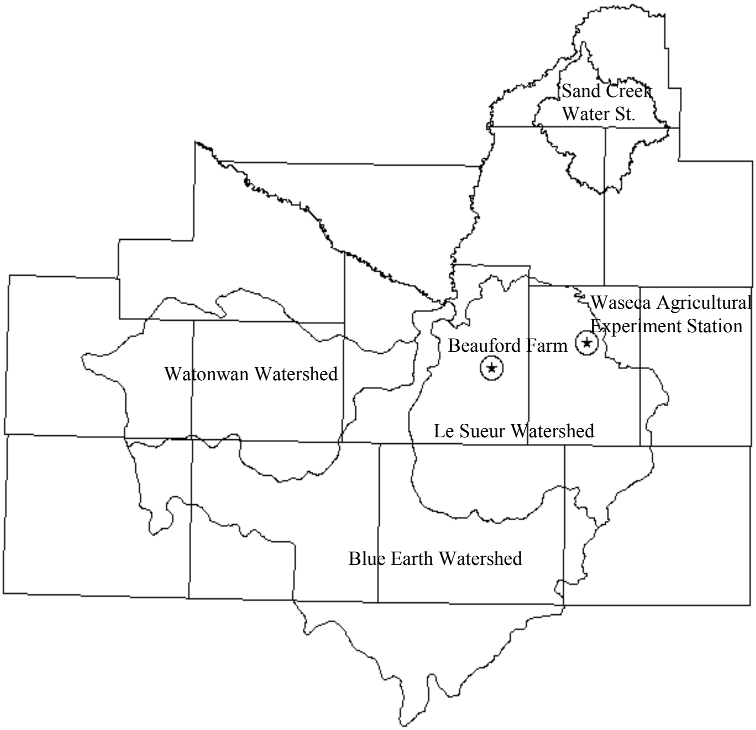
In the four sub-sections that follow, various approaches are described for estimating nitrate losses from agricultural landscapes in Minnesota. The scales discussed range from a few square meters to almost a million hectares. The examples include University plot-scale research, farm-scale research, minor watershed-scale modeling, and major watershed-scale nitrogen balances. The general location of each study is shown in Figure 1. For each scale, there are distinctly different approaches used, including the spatial resolution of input data, the method of estimating spatial variability, the method for flow routing, and the extent of mechanistic modeling versus mass balancing.

4.1 Nitrogen Losses at the Plot-Scale

At the plot-scale (m^2 resolution), experiments can be conducted to determine the leaching or drainage losses of nitrogen under conditions that are rigorously controlled and measured. As an example consider experimental measurements of nitrate in tile drainage water conducted at the Waseca Agricultural Experiment Station in Minnesota over a period of 13 years (Randall and Iragavarapu, 1995). Three plots of 13.5 m x 15.0 m in area were treated with 202 kg/ha nitrogen as a broadcast application in spring. Plots were planted to continuous corn (Zea mays L.) and subjected to conventional moldboard plowing. Tile drain spacing and depth were 27 m and 1.2 m, respectively. These plots were managed using University of Minnesota Best Management Practices, so nitrate losses may be considerably less than those observed on more poorly managed farm fields receiving similar rates of nitrogen fertilizer and/or manure.

A process based simulation model, the Agricultural Drainage and Pesticide Transport (ADAPT) model (Chung et al., 1992), was used to evaluate the relative impacts of rate of applied nitrogen fertilizer, and depth or spacing of tile drains, on nitrate losses in drainage water. The ADAPT model operates at a daily time step, and accounts for hydrology (precipitation, soil frost, snowmelt, evapotranspiration, infiltration, runoff, drainage, seepage), nutrient cycling (fertilizer dissolution, nitrification, denitrification, mineralization, immobilization, leaching, uptake), erosion , and crop growth.

Figure 1: A map showing locations of the Blue Earth, Le Sueur and Watonwan major watersheds, the Sand Creek minor watershed, the Beauford Farm, and the experimental plots at the Waseca Agricultural Experiment Station in southern Minnesota.



The ADAPT model climatic inputs include rainfall, temperature, wind speed, relative humidity, and solar radiation. Soil inputs include horizon thickness, porosity, moisture characteristic curve, organic matter content, saturated hydraulic conductivity, and texture. Management parameters needed by the model include rate, timing, and method of fertilizer application, dates and types of tillage, depth and spacing of tile drains, and dates of planting and harvesting. Other important input parameters include slope steepness, runoff curve number, crop rooting depth, and depth to impermeable layer.

Half of the experimental years were used to calibrate the model, the other half to validate the model (Davis, 1998). Four statistical procedures were used to evaluate the model performance during calibration and validation. These include i) the observed and predicted means and standard deviations, ii) the coefficient of determination, iii) the slope and intercept of a least squares regression between the predicted and observed values, and iv) the root mean square error. After validation, the model was used to evaluate nitrogen application rate (100, 125, 150, 175, 200, and 225 kg/ha), tile spacing (15, 27, 40, 80, 100, and 200 m), and tile depth (0.9, 1.2, and 1.5 m) using a continuous corn rotation with a 50-year climatic record for the site. Results for nitrate losses in drainage were ranked in decreasing order, and summarized using exceedance probabilities.

There was excellent agreement between observed and predicted nitrate losses in drainage during calibration and validation of the model (Davis, 1998). Observed annual nitrate losses during the 13 years of plot data collection were 41 kg/ha, while predicted losses were 55 kg/ha. The root mean square error was 0.85 during calibration, and 0.94 during validation.

For a tile drain spacing of 27 m and a tile drain depth of 1.2 m, regression analysis was used to show the average relationship between growing season precipitation and nitrate losses in drainage for different simulated rates of applied nitrogen fertilizer. In general, nitrate losses increased linearly with precipitation. The rate of increase depended on the rate applied, with faster increases (greater slopes for the regression lines) with higher rates of application. Once nitrogen application rates were increased above the recommended guidelines (125 kg/ha), the excess nitrogen was very susceptible to being lost in drainage water. At a nitrogen application rate of 150 kg/ha, the annual nitrate losses averaged about 10 kg/ha. For an application rate of 175 kg/ha, nitrate losses in drainage averaged about 23 kg/ha.

Instead of evaluating average behavior, we also evaluated the probability of a particular nitrate loss using exceedance probabilities. For any applied rate of fertilizer nitrogen, an exceedance probability curve was developed. An exceedance probability of 0.5 corresponds to the median behavior. An exceedance probability of 0.2 corresponds to rare climatic events that occur less than 20% of the time. An exceedance probability of 0.8 corresponds to frequent climatic events that occur 80% of the time. As an example of this approach, with an applied fertilizer rate of 175 kg/ha, the exceedance probabilities for a 10, 20, or 30 kg/ha nitrate loss were 0.87, 0.63, and 0.34, respectively. Clearly, if policy demanded that nitrate losses of 20 kg/ha did not occur greater than 50% of the time, an application rate of 175 kg/ha would not be allowed.

4.2 Nitrogen Losses at the Field-Scale

At the field-scale (32 ha), we measured nitrate losses through tile drainage for one year in a field located in the Beauford watershed in southern Minnesota (Mulla et al., 2001a). The field was divided into a steeper southern portion and a flatter northern portion. The southern portion has less subsurface tile drainage than the northern portion. Runoff and tile drainage for each portion were measured separately. Nitrogen fertilizer was applied uniformly to the field in fall at typical rates. Measured losses of nitrate in tile drainage were very low (< 1 kg/ha) due to a dry climatic pattern.

The losses of nitrate were simulated using a corn-soybean rotation and a soybean-corn rotation with a 50-yr climatic record. Input data included spatial variability in soil organic carbon (from soil sampling and kriging), topography (from elevation surveys), and soil physical properties (saturated hydraulic conductivity, moisture retention curves, porosity, texture) from soil databases. The site has landscapes ranging from 0-6% slope steepness, and soil organic carbon contents ranging from 2-5%. The field has seven mapped soil series, mostly finer textured loams, silt loams, silty clay loams, and clay loams, with hydrologic classes B, C, B/D, and C/D, where B indicates good internal drainage, C indicates somewhat poor drainage, and D indicates poor internal drainage. The field was subdivided into one hundred twenty eight 50 m x 50 m cells, and each cell was assigned input parameters based on available information. Subsurface drainage was routed from cell to cell based on available maps of subsurface tile drainage. Roughly 60% of the field is tile drained. Surface runoff was routed from cell to cell based on surface topography.

The 50-yr simulations were run using uniform application rates of 120, 140, and 160 kg/ha nitrogen fertilizer applied in the fall. The northern and southern portions of the field were kept separate, due to a topographic crest dividing the two portions. Exceedance probability curves were developed for each applied rate of fertilizer on each portion of the field. The median long-term rates of nitrate loss through tile drainage increased with application rate, with losses of 26, 28, and 31 kg/ha nitrate in the southern portion of the field at rates of 120, 140, and 160 kg/ha, respectively. Median losses of 33, 36, and 39 kg/ha were predicted on the northern portion of the field at applied rates of 120, 140, and 160 kg/ha, respectively.

As we shift from simulating nitrate losses at the plot-scale to losses at the field-scale, there is a similarity in methods used for the simulation, but a difference in the handling of input parameters. At the plot-scale we do not use information concerning spatial variability of soil properties or management inputs, and routing of flows is not necessary. At the field-scale, information about spatial variability and routing of flows becomes important.

4.3 Nitrogen Fluxes at the Minor Watershed-Scale

At the minor watershed-scale, we (Dalzell et al., 1999) used the ADAPT model to simulate flow and nitrate losses at the mouth of the Sand Creek watershed (651 km^2). Model inputs were obtained from various sources. The Minnesota State Climatology Office provided daily climatic information from 11 recording stations in and around the watershed. Daily results from these stations were averaged and used as input to the model. Information concerning fertilizer application rates and dates of planting and harvesting of crops was obtained from the Minnesota Agricultural Statistics Office. Water quality and flow monitoring data for three years (1994-1996) were available at the mouth of Sand Creek from the Twin Cities Metropolitan Council for Environmental Services. Monthly loadings of nitrate were estimated using monitoring data and a regression model based on daily flow. Soil physical properties were obtained from the Map Unit User File (MUUF) soils database associated with the State Soil Geographic (STATSGO) database maintained by the USDA-NRCS. Roughly 40% of the soils in

Sand Creek watershed are classified by the Soil Survey as having poor internal drainage. These soils are assumed to be artificially drained with subsurface tile drainage systems to reduce excess water and improve soil productivity. Landuse information (corn, soybean, grass, forest, urban areas, roads, open water) was obtained using LANDSAT remote sensing images and ground truth sampling at 85 locations from July, 1995. Only 63% of this watershed is in cultivated cropland, with substantial portions in grass and pasture, forest, and urban landuses. Crop residue cover was estimated using LANDSAT TM band 5 reflectance values from May, 1997, ground truth data from 85 locations, and a logistic regression equation. Slope steepness was estimated using a 30 m resolution digital elevation model for the watershed.

Hydrologically unique spatial data units were identified by overlaying GIS layers of slope, soil properties, landuse, and residue cover. Each resulting polygon is unique in terms of both model inputs and spatial location, and is referred to as a Hydrologic Response Unit (HRU). Running a separate model simulation for each HRU would be prohibitively time consuming, as there are over 50,000 HRUs for Sand Creek. Spatial data units (HRUs) were arranged into functional modeling units called Transformed Hydrologic Response Units (THRUs). The THRUs are distinct in terms of model input parameters, but include HRUs at different locations with identical input parameters. The ADAPT model is run for each THRU, a total of only 81 simulations. Output from each THRU is routed to the mouth of Sand Creek within one day, and the routing algorithm accounts for denitrification losses of nitrate beyond the edge-of-field.

The ADAPT model for Sand Creek predicted nitrate loads very accurately over the three year period from 1994 to 1997. The overall least squares regression line between predicted and observed nitrate had a coefficient of determination (r^2) of 0.78, with a slope of 1.59 tons/mo and an intercept of 1.37 tons/mo. Under baseline conditions (140 kg/ha of applied N fertilizer applied half in fall and half in spring), the loading to Sand Creek was 511 tons/yr (an average loss of 7.8 kg/ha over all types of landuses, including corn, soybeans, grass and alfalfa, forest, and urban areas. Alternative scenarios with a twenty percent increase or decrease in applied nitrogen fertilizer rate were also simulated. The model showed that a 20% reduction in N fertilizer gives a 6.8% reduction in nitrate loads in Sand Creek. A 20% increase in N fertilizer gives a 4.5% increase in nitrate loads for the watershed.

Modeling nitrate losses at the minor watershed involves the concept of transformed hydrologic response units (THRUs), as opposed to the use of simple landscape based hydrologic response units (HRUs) which are used at the field-scale. Also, there are differences in the spatial resolution of model input data. Spatial resolution of model input data is 30 m for digital elevation models at the minor watershed-scale, and 1 m for field topography. Spatial resolution of soil mapping units is at a scale of 1:5000 at the field-scale and 1:125,000 for the minor watershed-scale. Spatial resolution for crop planting and harvesting dates is at the scale of counties for the minor watershed modeling, in comparison to the scale of a single field. In addition to these differences, the combination of modeling and GIS database analysis becomes much more important at the minor watershed-scale than the modeling approach at the field-scale. Finally, routing algorithms at the minor watershed-scale account for in-stream denitrification losses, while routing algorithms at the field-scale do not.

4.4 Nitrogen Fluxes at the Major Watershed-Scale

Nitrogen fluxes have been monitored at the scale of major watersheds in the Minnesota River Basin for roughly 20 years (1974-1994). The Minnesota River Basin has twelve major watersheds, of which three watersheds generate two-thirds of the nitrate loads (Mulla and Mallawatantri, 1997) carried to the mouth of the Minnesota River (31,670 tons nitrate/yr out of a total of 50,270 tons/yr). These three watersheds, which cover only 20% of the area in the Minnesota River Basin (8,176 km^2), are the Blue Earth, Le Sueur, and Watonwan watersheds. According to Soil Survey information, about 55% of the soils in these watersheds are poorly drained, and have artificial tile drainage. Row crop agriculture accounts for 92% of the land uses, and cropland receives high rates of manure and nitrogen fertilizer, has high organic matter contents (>4%), and is in a humid climatic region with greater than 750 mm/yr of precipitation. All of these conditions contribute to excess nitrogen applied to cropland, which is then transported through tile drains to surface waters in large quantities.

We calculated the excess nitrogen applied to row crops in these three watersheds using a mass balance approach, and compared the excess nitrogen with the monitored nitrate loads carried by surface waters (Mulla et al., 2001b). We assume that most of the excess nitrogen (the exception being N loss by denitrification) is transported to surface waters after leaching to subsurface tile drains and discharge in surface ditches. Calculations of excess nitrogen were accomplished using a combination of nutrient recommendations from the University of Minnesota, agricultural statistical data on fertilizer applications and fertilized cropland area at the scale of counties, landuse datalayers for crop types, statistical surveys of farmers applying both manure and fertilizer to cropland, animal inventory data for each feedlot in the watersheds, a state permitting feedlot database, research data on manure N content by animal species and weight, research data on manure N losses for various types of animal confinement, storage, and application methods, and GIS analysis of these datalayers.

The first and most complicated step in this process was to estimate the rates of N applied to cropland from manure. To do this we estimated the manure produced by animals in each feedlot from the feedlot inventory based on species type, numbers, and weight distributions. This manure production was converted to nitrogen production using average manure N contents. Losses of manure N were estimated based on a weighted average of the losses occurring in various types of animal confinement and storage types, with the information about confinement and storage types obtained from the state permitting database for feedlots. The permit database also contained information about the area of land available for spreading of manure. Linear regression was used for each animal species to relate animal units to the acres of available land for manure spreading. Good linear relationships were found for the predominant animal species in this area, namely; hogs, beef, and dairy cattle. For each feedlot, the regression line for the dominant species was used to estimate the area of land available for manure spreading. Manure N losses during application were estimated based on a weighted average of application methods from the permit database. The remaining manure N was multiplied by 0.55 to give first year available N applied to the land. This gives the amount of manure N applied to the land which was potentially available for leaching or crop uptake.

The land available for manure spreading in any single year was estimated as a fraction of the total land available for manure spreading in multiple years based on farmer surveys. These surveys provided the crop mixes used by animal operations, and the percent of each crop receiving manure in any single year. Half of the cropland in this region is in corn, the other half is in soybeans. University of Minnesota fertilizer guidelines recommend 134 kg/ha nitrogen for corn, and no nitrogen for soybeans. Farmers typically apply manure to 25% of their corn land at rates of 65 kg/ha, and to 10% of their soybeans at rates of 55 kg/ha in any single year. The 25% of corn land receiving manure also receives commercial N fertilizer at rates of 161 kg/ha. The rate of manure N applied in a single year was estimated based on the amount of manure N applied to the land divided by the cropland area available for manure spreading in a single year.

Aside from manured cropland, there is also unmanured cropland to consider. Potentially manured corn land which does not receive manure in any single year (75% of potentially manured corn land) is fertilized with commercial N fertilizer at rates of 185 kg/ha. In addition, about 43% of the total cropland in the watersheds is corn land which never receives manure, and this land also receives N fertilizer at rates of 185 kg/ha. The remaining cropland (47% of the total cropland area) is soybeans which does not receive nitrogen fertilizer applications.

Excess N applications for the three major watersheds were obtained by subtracting the recommended amounts of N from the fertilizer and/or manure N applied to cropland (whether manured or unmanured). This approach gives 28,701 tons/yr of excess nitrogen applied within the Blue Earth, Le Sueur, and Watonwan major watersheds. This is about 59 kg/ha of excess applied N on corn acres, which means farmers are applying 43% more nitrogen than the University of Minnesota recommended rate of 134 kg/ha for corn. Of the excess N applied to cropland, 79% is from fertilizer, and only 21% is from manure.

There are major differences in the approaches for estimating N losses at the major watershed-scale versus the minor watershed-scale. At the major watershed-scale there is an increased reliance on GIS and statistical mass balancing techniques, and a reduced reliance on process based simulation. Process based modeling becomes unwieldy at the scale of major watersheds, and the uncertainties and coarse spatial resolution in input parameters do not justify a modeling approach.

4.5 Up-Scaling Nitrogen Losses

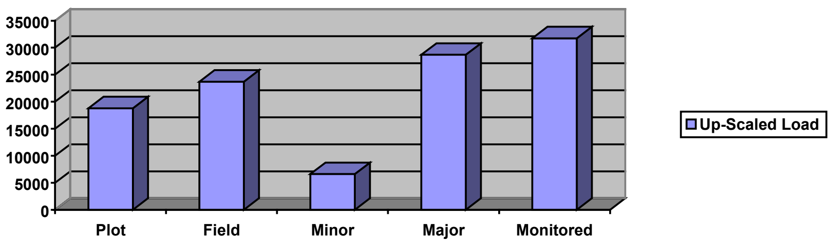
The nitrogen losses estimated at the plot, field, and minor watershed-scales can be used to independently estimate nitrogen losses at the major watershed-scale through up-scaling. The simplest approach to up-scaling is to simply multiply the N loss rates at each scale (for a scenario involving from 160-170 kg N/ha applied to corn) by the total area of the three watersheds in corn (408,810 ha). Another 408,810 ha is in unfertilized soybeans, which receives no fertilizer. Leaching losses from soybean land also have to be accounted for.

At the plot-scale, the long-term simulated N loss from a 175 kg/ha spring fertilizer application to corn was 23 kg/ha. At the field-scale, the long-term N simulated loss from a 160 kg/ha fall fertilizer application to corn was 35 kg/ha, while the long-term loss during the soybean portion of the rotation was 23 kg/ha. At the minor watershed-scale, the short-term N losses from 168 kg/ha applied half in fall and half in spring to corn were 8.1 kg/ha. Multiplying the watershed area by these losses gives total N losses of 18,805 tons N/yr based on the plot-scale modeling, 23,710 tons N/yr based on field-scale modeling, and 6,622 tons N/yr based on minor watershed-scale modeling.

As mentioned previously, long-term water quality monitoring shows that the Blue Earth, Le Sueur, and Watonwan major watersheds carry 31,741 tons nitrate/yr. We would not, however, expect tile drainage nitrate loads from these three watersheds to exactly equal 31,741 tons/yr, because sources such as atmospheric deposition, point source discharges, and interflow into drainage ditches should also contribute to the total nitrate load monitored. The up-scaled N leaching from the field-scale modeling (23,710 tons/yr) is, however, very similar to the monitored nitrate load carried by the three watersheds (Fig. 2). The excess nitrogen applications (28,701 tons/yr) estimated using mass balancing (detailed feedlot inventory and fertilizer sales data) without simulation modeling are also very similar to the monitored nitrate loads (Fig. 2).

The up-scaled nitrate losses from modeling at the plot- or minor watershed-scales are significantly lower than the monitored nitrate loads at the scale of major watersheds (Fig. 2). Plot-scale simulations are based on spring applied nitrogen, which is less susceptible to leaching than nitrogen applied in the fall at the field-scale study site. The up-scaled results from Sand Creek minor watershed are significantly lower than monitored loads because Sand creek watershed is significantly steeper and better drained than the landscapes and soils in the flatter and more poorly drained Blue Earth, Le Sueur, and Watonwan major watersheds. Also, Sand Creek has a significant portion of non-agricultural landuses, whereas the Blue Earth, Le Sueur, and Watonwan watersheds are overwhelmingly agricultural.

Figure 2: *Comparison between nitrate-N loads (tons/yr) at different scales after up-scaling.*



5. Conclusions

Process based simulation models are increasingly being used across a wide range of spatial scales to investigate nitrate losses to the environment. As spatial scale becomes coarser, model input data becomes less precise in terms of spatial resolution. At coarser spatial scales the importance of routing algorithms increases relative to finer spatial scales. Input data aggregation and classification becomes more important at coarser spatial scales than at finer spatial scales. Mass balancing approaches seem more appropriate at the coarsest spatial scale than mechanistic modeling.

The performance of spatial up-scaling techniques does not seem to depend as much on the extent of up-scaling as on the relative similarity between the smaller unit being up-scaled and the larger unit. In particular, the fraction of land which is tile drained and the rates and timing of fertilizer application must be similar. In this paper, best results for up-scaling to the major watershed-scale (10,000 km^2) were obtained from field-scale (32 ha) simulations, rather than from the minor watershed-scale (800 km^2) or plot-scale (m^2) simulations. This is because the field-scale conditions (soils, landscapes, climate, management) were more similar to the major watershed conditions than conditions at the plot or minor watershed-scale.

At the major watershed-scale, nitrate simulation modeling was not attempted due to issues of complexity and spatial variability of model input parameters. Instead, nitrate losses were estimated using a mass balance approach involving statistical databases and a GIS approach. Nitrate leaching potentials with the latter approach (28,701 tons/yr) compared satisfactorily with the up-scaled results from field-scale simulations (23,710 tons/yr), and with long-term water quality monitoring data (31,741 tons/yr).

References

Addiscott T. M., Smith, J., & Bradbury, N. (1995) Critical evaluation of models and their parameters, *J. Environ. Qual.*, 24, 803-807.

Chung, S.O., Ward, A. D., & Shalk, C. W. (1992) Evaluation of the hydrologic component of the ADAPT water table management model, *Trans. Am. Soc. Ag. Eng.*, 35, 571-579.

Davis, D. M. (1998) Simulation of tile drainage and nitrate loss from a clay loam soil, M.S. thesis, Univ. Minnesota.

Dalzell, B. J., Gowda, P. H., & Mulla, D. J. (1999) Predicting nonpoint source pollution for an agricultural watershed in southern Minnesota, Paper No. 99-2215*, Am. Soc. Ag. Eng.*

Jury, W. A. Gruber, J. (1989) A stochastic analysis of the influence of soil and climatic variability on the estimate of pesticide groundwater contamination potential, *Water Resour. Res.*, 25, 2465-2474.

Kirby, M. J., Imeson, A. C., Bergkamp, G., & Cammeraat, L. H. (1996) Scaling up processes and models from the field plot to the watershed and regional areas*, J. Soil Water Conserv.*, 51, 391-396.

Lark, R. M. Webster, R. (1999) Analysis and elucidation of soil variation using wavelets, *Eur. J. Soil Sci.*, 50, 185-206.

Loague, K., Bernknopf, R. L., Green, R. E., & Giambelluca, T. W. (1996) Uncertainty of groundwater vulnerability assessments for agricultural regions in Hawaii: Review, *J. Environ. Qual.*, 15, 15-32.

Mulla, D. J. Addiscott, T. M. (1999) Validation approaches for field-, basin-, and regional-scale water quality models, in: *Geophysical Mono. 108, Assessment of Non-Point Source Pollution in the Vadose Zone*, Corwin, D. L., Loague, K., & Ellsworth, T. R. (Eds.), Am. Geophys. Union, 63-78.

Mulla, D. J., Gowda, P. H., Koskinen, W. C., Khakural, B. R., Johnson, G., & Robert, P. C. (2001a) Modeling the effect of precision agriculture: Pesticide losses to surface waters, in: *Terrestrial Field Dissipation Studies: Purpose, Design, and Interpretation*, Arthur, E. L. (Ed.), Am. Chem. Soc., (In press).

Mulla, D. J., Birr, A. S., Randall, G. W., Moncrief, J. F., Schmitt, M., Sekely, A., & Kerre, E. (2001b) Impacts of animal agriculture on water quality, in: Technical Work Papers on Animal Agriculture, Johnson, G. (Ed.), Minnesota Planning Office.

Mulla, D. J. Mallawatantri, A. P. (1997) Minnesota river basin water quality overview, Minnesota Extension Serv., FO-7079-E, Univ. Minnesota.

Randall, G. W. Iragavarapu, T. K. (1995) Impact of long-term tilage systems for continuous corn on nitrate-leaching to tile drainage, *J. Environ. Qual*., 24, 360-366.

Root T. L. Schneider, S. H. (1995) Ecology and climate: Research strategies and implications, *Science*, 269, 334-341.

Shumway, R. H., Biggar, J. W., Morkoc, F., Bazza, M., & Nielsen, D. R. (1989) Time and frequency domain analysis of field observations, Soil Sci., 147, 286-298.

Wagenet, R. J. Hutson, J. L. (1996) Scale dependency of solute transport modeling/GIS applications, *J. Environ. Qual*., 25, 499-510.

Spatial Distribution of the Desert Locust, *Schistocerca gregaria,* in the Plains of the Red Sea Coast of Sudan During the Winter of 1999

Wopke van der Werf[1], Gebremedhin Woldewahid[1], Tsedeke Abate[2], Munir Butrous[3], Osman Abdalla[4], Abdel Moneim Khidir[4], Bashir Mustafa[4], Ibrahim Magzoub[4], Osman Abdin[4], Alfred Stein[1] & Arnold van Huis[1]
[1]Wageningen University, Department of Plant Sciences, P.O. Box 8031, 6700 EH Wageningen, The Netherlands; [2]United Nations, Food & Agricultural Organisation, P.O. Box 1867, Sana'a, Yemen; [3]United Nations, Food & Agricultural Organisation, P.O. Box 1117, Khartoum, Sudan; [4]Sudan Plant Protection Department, P.O. Box 14, Khartoum, Sudan

Abstract: A systematic survey for solitary desert locusts, *Schistocerca gregaria,* was carried out in a 90 km long and 15 km wide stretch in the plains of the Red Sea coast of Sudan between 7 December 1999 and 11 February 2000. The purpose was to determine the distribution of the solitary locusts and the relationship of their density to habitat factors. Samples were taken in a 5 km grid on 9 occasions. Solitary locusts were mostly found in crop lands, with millet and patches of the weed *Heliotropium* spp. growing on fine sandy soils with relatively good moisture availability particularly in the upper 0-15 cm soil layer. Much lower locust densities were observed in grazing land, on gravel areas with scrub vegetation, and on Red Sea bordering silt soils with a high salt content. The results suggest that vegetation and soil data may be used as indicators of areas where solitary locusts are most likely to be found, yielding potential savings in survey effort. The generality of these findings, and the usefulness of using such indicators in survey, will be further tested.

1. Introduction

The desert locust, *Schistocerca gregaria,* is a migratory insect that can cause substantial damage to food crops in most of Africa north of the equator and a large part of south-western Asia (Waloff, 1966) Locusts occur in two 'forms': a low density form (solitary phase) and a high density form (gregarious phase). These phases are morphologically, physiologically and behaviourally distinct (Uvarov, 1971). Phase change is under hormonal and pheromonal control and depends, among other factors, on locust density (Roffey and Popov, 1968; Roffey et al., 1970). Gregarious locusts form swarms taking them beyond the deserts into agricultural areas where they can cause rapid and devastating crop losses. Locust swarms are the ultimate outcome of a sequence of population dynamic processes that lead from low density "recession" populations, via outbreaks and upsurges to high density "gregarious" plague populations (Magor, 1994). The craft and science of locust management is to detect and control a locust population before it develops too far along this sequence of progressive gregarization, and thus prevent widespread damage.

Sampling and monitoring is an essential component of locust survey and control. Sampling is a difficult task because the potential expanses of land on which locusts may be breeding is enormous, and the distribution of locusts over habitats at a range of scales is difficult to characterise. A very basic practical question is therefore: "where to look for locusts?". Knowledge about the distribution of locusts in potential breeding areas is essential for the development of efficient survey and detection procedures.

2. Materials and methods & description of the study area

During the winter of 1999/2000 we periodically sampled desert locust at a range of scales in known breeding areas at the Red Sea coast of Sudan. At the largest scale, we took samples in a 5 by 5 km grid covering an area of approximately 90 by 15 km, with a range of habitats including *Panicum* grassland with heavy grazing, crop land with millet, wadis, gravel areas with scrub vegetation, and salt affected silt soil areas near the Red Sea. Samples of adult locusts were taken by walking 3 transects of 400 m length at each site, and counting the number of locusts "flushed" (chased away) on a 1 m wide swath. At the lowest scale, the distribution of immature locusts 'hoppers' was determined at a 50 x 50 m grid resolution in *Heliotropium* vegetations. Hoppers were counted by random placing of quadrats with sides of fine wire mesh. Most of our data concern solitary adult locusts, but at the end of the season, some hot spots of gregarious hoppers were identified.

Soils in the study area are stratified from west to east, with gravel soils near the mountains in the west, coarse sands further east towards the sea, fine sand still further east, and silt soils near the Red Sea. Water arrives in the area in two forms: as direct rainfall, and through run-on from the mountains, via the "khors". These riverbeds are only wet after rains in the mountains. They diverge and spread out extensively over the sand and silt soils near the sea. Water availability is related to soil type, with the highest moisture availability in the fine sand and silt soils, and the lowest in the gravel terrains near the mountains. Vegetation varies according to soil and moisture conditions. Vegetation analyses indicate that five main types of vegetation may be distinguished (Table 1): "*Suaeda* land" (mostly on silt soils near the sea), "*Prosopis* land" (locally called "Mesquite", a recently introduced plant to rehabilitate the desert area, mostly on sand or gravel soils in highly disturbed areas near the main highway and settlements areas), "Crop land" (mostly on fine sand near khors), and "*Panicum* grazing land" (mostly on course sand) and "Scrub" (on gravel soils).

Of the 60 sample locations in the 5 x 5 km grid, seven were in "*Suaeda* land", 10 in "Crop land" with *Heliotropium* spp., 17 in "*Panicum* grazing land", and 26 in "Scrub". "*Prosopis* land" was not sampled in the winter of 1999/2000, but results in the winter of 2000/2001 indicate that the locust situation in them is similar to that in "*Suaeda* land". Eyeball estimates of the area proportions of the five vegetation types are included in Table 1.

During the winter of 1999/2000, water came mostly in the form of direct rainfall, and very little if any arrived through the khors. The amount of rainfall was 22 mm in October, 11 mm in November, and slightly more than 100 mm during December. This allowed a good biomass development of annual plants, especially on the fine sandy soils, where millet was sown in November. Less than 10 mm of rain fell in January, and no rain at all in February and March. As a result, the vegetation of annual plants that had developed in December and January had completely dried out by the end of March, and did not sustain locust development anymore.

Table 1: *Characterisation of the sampled area, according to five vegetation types*

	Suaeda	**Crop land**	***Panicum***	**Scrub**	***Prosopis***
Location	East, near the sea	West of *Suaeda*	West of Crop land	West, near the mountains	Along the highway near settlement area
Total area (%)[1]	25	1	23	50	1/10
Dominant soil	Silt	Fine sand	Coarse sand	Gravel	Gravel & sand
Water availability (relative)	Good, but physiologically unavailable	Good	Medium	Low	Medium
Dominant perennial plant species	*Suaeda monoica;*		*Panicum turgidum*	*Acacia*	*Prosopis chilensis;*
Dominant annual plant species	*Cenchrus* , *Cyprus, Tribulus*, and others	Millet, *Heliotropium*, *Tribulus*	*Tribulus, Cenchrus,.* many grasses and dicotyledons	*Zygophyllum simplex, Cassia senna*	Annual plants mostly suppressed
Density of locusts (relative)	Low	High	Low	Low	Low

3. Locust observations

During the first sample (7 December), the density of locusts, averaged over all sample sites, was 2 ±1 (SEM) locusts/ha. The density increased during December, and the highest density of all 9 sample dates was observed on 22 December: 37 ±12 locusts/ha. Thereafter, the density gradually declined to 12 ± 7 locusts/ha on 9 February. As no hoppers were found in November or December, it is likely that the increase of adult locust density during December was due to immigration, while the gradual and slow decline during January and February may have been due to a combination of emigration and attrition. It was common to find dead locusts in January and February.

Hoppers were found for the first time in mid February, and exclusively in croplands. Most hoppers were feeding on the weed *Heliotropium* spp.. Some were also found feeding on millet. Other plants were less commonly observed being fed upon. Hopper densities decreased gradually until they were virtually absent by mid March. Gregarious hoppers were found in one of the croplands, near Khor Ashat on February 23. These hoppers developed into adults at the end of February and early March, but a substantial proportion died before reaching maturity. By the end of March, the crops were harvested, and very low locust numbers were found in the crop lands

The relationship between locusts and the environment was strikingly clear and straightforward. Locusts were only found in substantial numbers in crop land. For instance, at the end of December, when the densities were highest, the density of locusts was 148 ± 62 locusts/ha in crop land (with *Heliotropium* spp.), 21 ±8 locusts/ha in *Panicum* land, 32

---

[1] These eyeballing estimates of area are preliminary, and will be improved.

±13 locusts/ha in *Suaeda* areas, and 6 ±3 locusts/ha at scrub sites. As noted in Table 1, there are strong correlations between vegetation, soil type and water availability. Thus, high solitary locust densities in this area, in the winter of 1999/2000, were associated with the fine sandy soils with the relatively good water availability where the crops and *Heliotropium* were growing. Maxwell-Darling (1936) also found the greatest number of solitary locusts among rain grown millet at the coastal plains of the Red Sea coast in Sudan. Young hoppers seemed to prefer the weed *Heliotropium undulatum* in cultivated fields, acting as a concentration factor. In such fields incipient outbreaks were observed in two seasons. However, it was argued that the cultivation in itself may not have been of primary importance, but that the loose sandy soil was a decisive factor.

4. Discussion

When this research was initiated, one of the working hypotheses to be tested was that predictions of locusts at distant locations might be made on the basis of spatial correlations with data obtained at sampled sites. Results show that habitat factors are of overriding importance, and that nearby sites with different vegetations have very dissimilar locust densities, whereas distant sites with similar vegetations have similar locust densities. This finding indicates that locust survey might be organised efficiently by surveying those vegetations that are likely to sustain locusts, rather irrespective of their location, rather than surveying sites in a more or less regular grid over the region, thereby aiming at obtaining estimates of densities at intermediate locations by a method of spatial estimation, e.g. kriging. It will thereby be crucial to determine what environmental factors make sites suitable for locusts. The results of this study in the coastal regions of Sudan, indicate that the crop lands, with millet and *Heliotropium* spp. as most abundant plant species (which are both suitable for locust development), are the hot spots where solitary locusts might be found multiplying, and where survey effort should be focused. Work is underway to further validate these results.

Acknowledgements

The research reported here is carried out cooperatively by Wageningen University, FAO, and the Plant Protection Department of Sudan. It is supported financially by the Dutch Ministry of Foreign Affairs and is linked to the FAO program EMPRES (EMergency PREvention Systems for Transboundary Pests and Diseases).

References

Magor, J.I. 1994. Desert Locust population dynamics. In: Van Huis. A (ed.), *Desert locust control with existing techniques: an evaluation of strategies*. Proceeding of the seminar held in Wageningen, The Netherlands 6-11 December, 1993, Wageningen Agricultural University, 31-54 pp.

Maxwell-Darling, R.C., 1936. *The outbreak centres of Schistocerca gregaria*, Forsk., on the Red Sea coast of the Sudan. Bull. ent. Res. 27: 37-66.

Roffey, J., Popov, G.B. 1968. Environmental and behavioral processes in a Desert Locust outbreak. *Nature*, Lond. 219, 446-680.

Roffey, J., Popov, G.B. and Hemming, C. F. 1970. *Outbreaks and recession populations of the Desert Locust*. Bull. ent. Res. 59, 675-680.

Uvarov, B.P. 1977. *Grasshoppers and locusts. A handbook of general acridology*. Vol. 1&2. London, Center for Overseas Pest Research.

Waloff, Z., 1966. The upsurges and recessions of the Desert Locust plague: a historical survey. *Anti-Locust Memoir No. 8*, 111 pp.

Integration of Environmental Aspects into Agricultural Policy: the Case of the European Union[1]

García Azcárate Tomás
Directorate for Economic Analysis and Evaluation - Agricultural and Rural Development - European Commission
Rue de la Loi 130 7/137 B-1049 Bruxelles, Belgium
Tel: +32 2 2953317 Fax: +32 2 2963987, e-mail: tomas.garcia-azcarate@cec.eu.int

Scheele Martin
Institute of Agricultural Economics – University of Bonn
e-mail: martin.scheele@cec.eu.int

1. Introduction

This paper gives an overview on the Common Agricultural Policy (CAP) from the perspective of the integration of environmental aspects. It starts with a description of recent developments of the policy framework, followed by a brief reflection follows concerning the implications of certain CAP elements on various topics of the integration strategy.

2. The Policy Framework of the Common Agricultural Policy

As regards the domestic priorities for EU agriculture, the centre of attention is the improvement of competitiveness and using the opportunities provided by expanding world markets. Important concerns in this respect are continued structural adjustments in primary production, high standards of food safety and quality, efficient and environmentally sound technologies, and efficient marketing structures.
The CAP reform under Agenda 2000 represents a further important and wide ranging reform of the Common Agricultural Policy. This is reflected in the range of sectors involved: cereals, oilseeds, beef, milk and wine. If the market organisations for fruit and vegetables, tobacco and olive oil are taken into account (which were recently reformed), almost the total value of the Union's agricultural output is affected by these reforms.
Agenda 2000 sets the architecture for European agricultural policy for the coming years and consolidates the shift that took place during the 1992 reform. To understand the scale of this change it is necessary to place it in a historical context.

---

[1] The view presented in this contribution are those of the author and do not necessary represent those of the Commission.

2.1. The Early Days of the CAP

Since its inception in the 1960s and throughout the 1970s and 1980s market price support was the dominant form of support for most agricultural products in the European Community.

During the 1970s and 1980s productivity and production rose. It became increasingly difficult to avoid the build-up of surpluses. It therefore became necessary to apply supply control mechanisms to curb overproduction such as the introduction of the milk quotas in 1984 and the stabilisers in 1988 (policies which provided an automatic link between price reductions and production when the latter exceeded certain thresholds).

2.2. The 1992 Reform

During the 1980s the realisation grew that the disadvantages of price support, i.e. support linked to the volume of production, were increasingly outweighing any advantages. The need for administratively complicated supply controls constrained the efficient development of agriculture. Budgetary expenditure on intervention stocks and export subsidies continued to increase (despite supply controls).

The demand for agricultural commodities was being eroded by high prices, excessive production intensity was damaging the environment, and the policy was not particularly effective in supporting farm income. This prepared the ground for a major policy shift in the early nineties leading to the 1992 CAP reform decision based on the following principles:

- a substantial reduction in the prices of a number of key agricultural products (cereals and beef) to make them more competitive both within the Community and on world markets;
- compensation for the price cuts in the form of hectare or headage payments with upper limits based on historic data and not linked to output;
- implementation of measures to limit the use of the factors of production (set aside of arable land and stocking rate criteria for livestock);
- introduction of accompanying measures to the market reforms promoting environmentally friendly farming, afforestation and early retirement.

The 1992 decisions also included a major reform of the tobacco sector (ending of intervention and limitation of support to quotas per variety), the ending of price support for oilseeds and protein crops with a shift to area payments and a reduction in the intervention price for butter.

2.3. CAP Reforms under Agenda 2000

The Agenda 2000 embodies a two pronged approach, continued market reform and development of a second pillar to the CAP englobing structural, environmental and rural policy. The main elements of reforms in both parts of the CAP are outlined in the following sections.

2.3.1. Reforms of Market Organisation under Agenda 2000

The reforms under Agenda 2000 followed the insight that market policy cannot continue to assume the function of permanent intervention. Therefore, action was undertaken to limiting the role of intervention to that of a safety net in cases of substantial price drops. If the period of the reform since 1992 is taken as a whole, support prices for cereals will have fallen by 45% and beef by 35%.

The reforms cover a wide range of areas from legislative and administrative simplification, through the integration of environmental concerns to changes in budgetary management. The Commission's proposals were based on an analysis and a forecast of market developments, in the EU and world-wide. In particular for cereals and beef, these analyses showed clear risks of increasing production surpluses post 2000.
As regards key commodities, the main changes can be summarised as follows:

– For cereals support prices have been cut by 15%. This reduction will be phased in from 2000. These cuts are partially compensated. Oilseeds payments are aligned on those of cereals. Compulsory set-aside is retained with a basic rate of 10%.
– The support level for beef has been reduced by 20%. Public intervention will be replaced by a system of private storage. Intervention buying plays only a role as a low-level safety net and as a possible ad hoc measure in the case of a crisis. These changes will be phased in from 2000. The system of premia has been modified to place more emphasis on regional differences and extensification.
– The support price for milk, SMP and butter will be reduced by 15%. This will be phased in from 2005 with a 1.5% increase in quotas. Five member states will receive special quota increases from 2000. The future of the regime will be reviewed in 2003 with the aim of phasing out the current arrangements.

Direct payments have played an important role in encouraging European farmers to adapt to new conditions. Together with an integrated rural development policy they account for 80 % of total CAP expenditures, whereas only 20 % are foreseen for market interventions.
Reflecting the increasing weight of direct payments a Regulation on Common Rules has been established. Among other aspects, this regulation includes provisions on modulation and cross-compliance both being optional on the side of Member States.

– Under the Article on modulation, Member States are authorised to reduce direct payments up to 20 % following the application of criteria such as low labour input, overall prosperity, and total amounts of payments granted.
– The article on "cross-compliance" authorises Member States to cut payments as a penalty for not complying with environmental legislation or specific environmental requirements established on the basis of the named regulation.
–

2.3.2. Building the second Pillar of the CAP: Rural Development Policies

The main fields of intervention of the newly established rural development policy can be broadly summarised under three headings:
– Measures to help strengthen the competitiveness and viability of the agricultural sector;
– Measures to improve the living conditions and economic opportunities in rural areas, particularly for those communities closest to changes in agricultural structures;
– Measures to promote good environmental practices as well as the provision of services linked to the maintenance of habitats, biodiversity and landscape.

Promoting Competitiveness and Viability

One of the most important measures for the improvement of the competitiveness of holdings is support for investment. This helps farmers to modernise buildings and equipment, improve the quality of management and develop alternative production potential on the farm, particularly in higher value added products. Investments to improve veterinary and phytosanitary hygiene, reduce environmental problems and promote animal welfare are also supported.

The conditions for EU support have been clarified and simplified. Part-time farmers are now fully eligible for support. Specific sectoral restrictions have been abolished. The general principle remains that investment is only possible in sectors where additional production has clear market outlets. But this is now applied on a programme by programme basis.

A second set of measures to improve competitiveness and develop new markets focus on the improvement of processing and marketing structures for agricultural products. Here again, there is scope to help farmers improve their orientation to consumer expectations particularly as regards quality, safety and traditional/organic production methods.

Basic and advanced vocational training is available to farmers who wish to improve their expertise in business management, technology, or new production processes. This can help enhance productivity, production quality and, indeed, the environmental performance of their holding. The possibility of vocational training has been extended to all persons active in the farming and forestry sector, and covers also forest management and general further education.

The improvement of the age structure of the agricultural labour force can be supported further by early retirement programmes and the helping the establishment of younger farmers. These measures have been simplified and made more attractive.

A range of measures can be undertaken to improve land, to consolidate holdings, to manage water resources intended for agriculture, or to support infrastructure associated with the agricultural development. These are also intended to help strengthen the agricultural competitiveness.

Improving living conditions and economic opportunities for rural communities

The second group of rural development measures concerns above all the improvement of living conditions as well as economic opportunities for rural communities outside the core area of agricultural production and marketing. Support is therefore available for the diversification of economic activities, both on and off-farm, ranging from increasing the range of local products to the development of craft activities and agri-tourism.

Forestry is recognised as an important part of development in rural areas development and is supported. In addition to traditional investment grants for modernisation and rationalisation, a new aid for the sustainable management of forests and forestry resources has been introduced. Where forests play a special role in environmental protection, as a buffer zone, or indeed protection from natural disasters such as fires, support is once again available.

An improvement of the economic conditions and the quality of life in rural areas is also to be achieved by a better access to basic services and infrastructure facilities. Appropriate measures can also be taken for the redevelopment and renewal of villages and the maintenance of rural cultural heritage.

As mentioned earlier, it is essential to ensure that in those areas eligible for support in the form of regional development programmes (the so-called Objective 1 and 2), rural development and regional development measures work effectively together.

Promoting Environmental Objectives

The third group of measures concerns environmental objectives and the preservation of the countryside. Behind this lies the goal of sustainable agriculture and forestry, which reflects the broader concern of sustainable development in rural areas.

Sustainable agriculture means ensuring that the benefits of natural resources and Europe's unique environmental heritage are available to future generations. Addressing the environmental dimension of the CAP includes at one level measures reducing the negative environmental impact of agriculture (e.g. investment aid for environmentally sound production methods, beef extensification premia). On the other, measures aim to secure the positive effects of farming activities on maintaining the rich rural landscape and biodiversity.

The core of the Community's agri-environmental strategy within the CAP is targeted agri-environmental measures which reward farmers for environmental services in rural areas, over and above good agricultural practice and environmental legislation. The inclusion of such measures into all rural development programmes implemented by Member States is compulsory.

Lying close to questions of viability, compensatory allowances are granted to less favoured areas with a view to compensate for natural disadvantages and, thereby, to ensure continued farming in those areas. This measure has an important environmental dimension in as far as farming plays an important role in providing non-commodity outputs such as landscape, agri-tourism or other rural amenities. Beyond this, compensatory allowances can also be granted to farms operating in Natura 2000 areas and subject to area-specific environmental constraints. An important additional change has been to shift from headage payments to area payments.

2.4. Sustainability: A Key Motive of the CAP Reform under Agenda 2000

Looking at the Agenda 2000 reforms from the perspective of sustainability, it can be summarised as follows:

- As regards the *economic dimension* of sustainability, a first-order condition was the shift from permanent *intervention* towards intervention as *safety net*. This makes EU commodities internationally more competitive. The changes are complemented by *rural development policies* becoming the second pillar of the CAP. Rural development programmes include measures improving the *competitiveness* of EU agriculture (investment aid, measures improving marketing and processing, training etc.)
- As regards the *social dimension*, disruptive pressures on the farm sector resulting from cuts in support prices are compensated through direct payments. The *social viability* of both rural areas and the farm sector is supported *by rural development measures*. An important concept in this respect is that of "*diversification*" of income source for farm households.
- As regards the *environmental dimension*, the general context is given by the Common Rules Regulation which obliges Member States to undertake appropriate environmental measures. In fulfilling their obligation, Member States have several

options at their disposal: agri-environmental measures, environmental legislation, and specific environmental requirements. The latter two options can be enforced by cutting direct payments granted under the Common Market Organisations. Further environmental elements not explicitly mentioned in the Common Rules Regulation are the beef extensification premia and payments for Less Favoured Areas (which can be used to finance the implementation of Natura 2000).

2.5. Strategy Papers on "Sustainable Agricultural and Rural Development"

Sustainable development has become a main motive of the more recent developments of the CAP. The concept and the directions pursued in this respect refer to Agenda 21 and, in particular, chapter 14 on "Sustainable Agriculture and Rural Development". Annex 1 to this document includes a list of references to the concept of sustainability as taken from strategy papers and legal texts.

The key strategy paper has been "Agenda 2000: For a Stronger and wider Union" which outlined in its agricultural chapter the directions of the CAP reform and the need for promoting sustainable rural development. Agenda 2000 was adopted in March 1999 by the European Council in Berlin.

Following the adoption of Agenda 2000, the Agriculture Council endorsed in October 1999 an *Integration Strategy* which was presented in December 1999 to the Helsinki European Council, in line with the overall integration process started by the European Council at Cardiff in June 1998.

The integration strategy of the Agriculture Council follows the directions established in February 1999 by the Commission *Communication "Directions towards sustainable agriculture"* COM (1999) 22, which had developed the environmental context for the Agenda 2000 proposals and underlined the need for a continuous process of integration and monitoring of progress.

In January 2000, the Commission presented its Communication *"Indicators for the Integration of Environmental Concerns into the Common Agricultural Policy"* COM (2000) 20, following a request of the Agricultural Council. This communication presents a framework of agri-environmental indicators with a view to provide the means for monitoring the implementation of the integration strategy and improving transparency, accountability, and evaluation.

In discussing COM (2000) 20, the Agricultural Council underlined that it is important to broaden the approach and to fully cover sustainable development, which includes the integration of the economic and social dimensions of sustainability. The request is being taken up in a currently prepared Commission Working Document. This working document presents a *"A Framework for Indicators for the Economic and Social Dimensions of Sustainable Agriculture and Rural Development"*. It is foreseen to present this paper at the Gothenburg Summit in June 2001.

3. Links between the CAP and some Key Issues

3.1. Impact of the reduction of production

Using farming practices constituting a risk for the environment is neither a direct result of the CAP, nor is it automatically associated, as often claimed, with certain types of "industrialised" agriculture. It is a result of farmers seeking to reduce production costs (e.g. by using cheap feeding stuff) or increasing yield. As regards the most actual issue of BSE, cases have been observed in all size categories and types of dairy farms. For dairy, the so-called "industrial agriculture" is anyway not sufficiently dominant to be responsible for major similar considerations to apply to using pesticides. Whereas it is true that price support has the effect of encouraging higher levels of pesticides, data from countries such as New Zealand which underwent a strict policy of liberalisation, show that cuts in price support are reflected in lower levels of inputs used. However, as time goes on, technical developments (high yield varieties requiring higher pesticide levels) and - as has been the case in the EU after the 1992 CAP reform - favourable market price developments can take over again the role as a driving force for increasing application levels.

It should be noted in this context that efficient and modern production methods are by no means bad or dangerous as such. They help farmers to contribute to a rational use of scarce resources and to stay in business against an ever-increasing competition from third countries. Of course, where cost-reducing practices prove unsafe for consumers or the environment, they have to be banned, as it was the case with feeding meat and bone meal to ruminants.

Nevertheless, indirect effects can be triggered by certain CAP measures. This includes the incentives that price support might provide for the intensification of farming, e.g. in terms of increasing pesticide use. Intensification effects might also result from certain investment aids.

Such risks have been significantly reduced by Agenda 2000 which, on the one hand, cuts price support (taking away the incentive for excessive input use) and, on the other, makes support under rural development programmes conditional on the "*compliance with minimum standards regarding the environment, hygiene and animal welfare requirements"*.

3.2. Climate Change and Clean Energy

Within the EU, agriculture is a major source of both CH_4 and N_2O, accounting for some 50% of N_2O and 40% of CH_4 emissions in 1990. Between 1990 and 1995, agricultural emissions of both *gases* fell by about 6%, due principally to changes in agricultural practices and production following the reform of the Common Agricultural Policy (CAP).

There are three main sources of emissions of greenhouse gas emissions from agriculture (approximate *contribution* in CO2 equivalent in 1995).

- N_2O emissions from soils (49%);
- CH_4 emissions from enteric fermentation (35%);
- CH_4 and N_2O emissions from manure management (15%).
- Other (1%).

Soil emissions of N_2O come from the application of both mineral N fertilisers and organic N in animal manure. Enteric fermentation is the anaerobic fermentation of polysaccharides and other components of animal feeds in the gut of ruminant animals (the rumen) by micro-organisms. Food enters the rumen where it is fermented to volatile fatty acids (VFA), carbon dioxide and methane. Both methane and nitrous oxide can be emitted from manure. Methane emissions depend on the animal's diet, manure management and the climate.

The 1992 CAP reforms have already had an impact on a number of parameters that affect these emissions. On the livestock side there have been changes in types and number of livestock, and an enhancement of the trend towards improved livestock productivity. On the arable side, the shift from production based support mechanisms to direct area based payments, has led to pressure to optimise the use of inputs economically, and a general reduction in fertiliser use. Agenda 2000 has reinforced these trends.

The projections of baseline emissions for the agricultural sector are made on the basis of projections of livestock numbers. Based on the implementation of Agenda 2000 policies, the projected baseline trend is a reduction of 7% in methane by 2010 and 11% for N_2O.

Underlying the change in emissions from livestock (enteric fermentation and manure) is a reduction in the number of cattle and an increase in the output of pig and poultry products. Production in the cattle sector is constrained by supply management with increasing efficiency providing a reduction in emissions. A relaxation of this supply control would halt or even reverse this downward trend, depending on its magnitude.

Another link between agriculture and climate issues is that of bio-fuels. The environmental merit of biomass is founded on its carbon cycle. Biomass emits in combustion not more CO_2 than assimilated during growth. This requires, however, that the energy-balance of the process as a whole remains positive (i.e. energy use for tractors, transport, fertilisers, pesticides, etc.). This condition can in general be assumed as given. The CAP includes only few elements favouring renewable energies. Rural Development Programmes include investment aid also applicable to technologies making use of bio-mass. Furthermore, they include aid for planting fast growing trees for combustion. Apart from this, farmers get permission to grow renewable resources on set-aside land.

Given the still existing gap between the price for fossil fuels and biofuels, there was some reluctance to opening a line new of farm subsidies in favour of biofuels. It is in this perspective that the Communication "Directions towards Sustainable Agriculture" (COM (1999) 22) stated that "*developing the non-food sector would need to be combined with appropriate fiscal measures.*

3.3. Natural Resources

Agriculture can have both negative and positive effects on the quality of natural resources.

- The negative environmental effects from agriculture on natural resources include the pollution of soils and water with nitrate (nitrogen balance surplus) and pesticides. Main concerns in this respect are the quality of drinking water, the eutrophication of surface water, and the implication of adverse chemical reactions in soils for erosion. Negative impacts of agriculture include other drivers for erosion such as inappropriate rotations, soil compaction, and overgrazing.
- Positive impacts of agriculture reflect the antropogenic origin of many high nature value landscapes. The ecological value of such landscapes depends on the

continuation of appropriate types of land use. A deterioration of the economic basis of farming would, therefore, lead to the abandonment or marginalisation of land use, in particular in naturally handicapped regions. The abandonment or marginalisation of land use would not only threaten the cultural value of the European landscapes. In many cases, it would be accompanied by soil erosion and loss of bio-diversity.

Negative environmental effects are regulated through a range of general environmental legislation (drinking water, ground water, surface water, chemicals, organic pollutants etc.). A sector-specific approach has been applied in the form of the EU *Directive on Nitrate Pollution from Agricultural Sources (91/676/EEC)*. The objectives of this directive are twofold: reducing water pollution by nitrates from agricultural sources and preventing further such pollution.

As regards the CAP and the efforts to integrate environmental requirements, the respect of such environmental legislation is supported through two main mechanisms:

- The respect of minimum environmental requirements is a condition for receiving support under the Rural Development Programmes. This rule holds also for agri-environmental measures which are not applicable to matters regulated by mandatory environmental legislation. The respect of the latter is considered to form part of "good farming practice" and agri-environmental measures grant payments only for commitments going beyond the "reference level" of good farming practice.
- In addition to this, Member States can make use of the option of linking direct payments granted under the Common Market Organisation to the compliance with mandatory or specifically defined environmental requirements (cross compliance).

Whereas measures aiming to avoid negative effects of agriculture normally shift the compliance costs to the farmers, a different approach is pursued where farmers - through engaging privately owned factors of production (e.g. land, labour, machinery) - provide environmental services. This approach is in particular applied to the maintenance of high nature value landscapes and biodiversity as well as to applying precautionary measures concerning the preservation of natural resources (water, soils).

The key environmental approach in this context is the agri-environmental measures which remunerate farmers on a voluntary, contractual basis for providing environmental services. Member States have to include agri-environmental measures into their rural development programmes. Another rural development measure to be mentioned here is the LFA scheme which promotes the continuation of farming and, thereby, environmentally relevant land management. A specific type of compensatory allowances can support the implementation of area-specific environmental policies (e.g. Natura 2000).

Whereas the integration of the environment into the CAP is a key requirement, it will not solve all environmental problems. Often, problems arise due to the non-application of existing environmental legislation (e.g. nitrate directive, FFH) and as a result of technical and economic development which exists independently from the CAP. Therefore both environmental integration and a strict application of environmental policies have their own responsibilities and must complement each other in meaningful way.

As regards the future perspectives, changes in agricultural structures resulting from increasing competitive pressures from global markets are expected to lead to a dichotomy in the development of land use. An increasing concentration of production in

regions with a good infrastructure and a high productive capacity on the one side and, on the other, marginalisation of the land use in areas disadvantaged by nature would be a realistic scenario that could have considerable negative environment implications.
Where the environmental functions of sustainable agriculture is no more ensured on the basis of private economic interest, more efforts would have to be undertaken with respect to providing such public goods through targeted agri-environmental measures.

3.4. Land Use

In quantitative terms, the bigger proportion of the land used is rural areas. Strengthening the economic viability of rural areas is the very basis for providing the means for preserving their social and environmental functions. These include important balance functions for overloaded centres of dense development, particularly through the preservation of a region's ecological integrity, as buffer zones, and recreation areas.
The significance of agriculture for Sustainable Rural Development is to a large degree based on agriculture being the most important user of land. Consequently agriculture experiences the confrontation with manifold interests relating to land.

- A key concern is a state of the rural environment and landscapes that correspond to society's cultural and esthetical values. This is not only a matter of the primary objective of "resource protection". It refers also to a series of secondary objectives such as the attractiveness of rural areas as a place to live in and for leisure activities and tourism.
- Another concern is the cultural identity of a certain region that is closely linked to farm structures and architecture, pattern of land use, local food processing, and handcrafts. Again, agriculture can assume functions beyond the narrow role of food production having important secondary economic effects (tourism, increasing the attractiveness of rural areas as a living space).
- As regards the narrowly defined economic functions of the farm sector, one can state that agriculture is to a diminishing degree a determining factor of the rural economy. In view of further increases of productivity within primary production, important perspectives for farm households with respect to employment and job creation would lie in the diversification of income sources (processing at primary stage, marketing of local products, rural tourism, rural handicraft, decentralised energy supply, local service capacities, environmental services etc.).

Many of the valued features of agriculture and rural areas do not come by default but require appropriate policies. The key to this is an integrated rural development policy, the second pillar of the CAP, which operates across the Union and accompanies the changes brought about by changes of market policies. The key objective of Rural Development Policies is to strengthen the competitiveness and viability of rural areas.
The concrete Rural Development Programmes may include:

- Investment in farm holdings;
- Young farmers' aid
- Pre-retirement
- Training
- Investment in the processing and marketing of produce
- Compensatory allowances in Less Favoured Areas
- Support for sustainable forestry

- Agri-environmental programmes
- Support for rural infrastructure, village renewal, tourism and craft activities

These measures aim to enable local and regional actors in rural areas to (i) respond to new demands for services in the environmental, tourism, cultural and amenity sectors, (ii) develop more diversified and higher value added agricultural and food production which is competitive internationally, and (iii) ensure balanced territorial development as regards the distribution of employment opportunity and economic activity.

4. Data Availability and Data Needs

The availability of data relating to sustainable agriculture and rural development is fairly good with respect to the economic dimension of primary production as well as farm structures.
A comprehensive data set exists on farm income (Farm Accountancy Data Network FADN), although limited to income from farming activities, which makes it difficult to draw conclusions on the social situation of farmers and farm families.
A general limitation of economic and structural data is the partly high level of aggregation (sectoral, spatial), which tends to hide the often significant differences between regions.
As regards environmentally relevant data relating to land (e.g. land use, land cover, input use, stocking density, semi-natural habitats, etc.), information is limited. Most of the data on input use is expressed in monetary terms while being insufficiently differentiated with respect to the different substances.
As regards the processing of environmentally relevant data much progress has been made by establishing a framework on *Agri-environmental Indicators* and a framework for *"Indicators for the Economic and Social Dimensions of Sustainable Agriculture and Rural Development"*.
With a view to filling this framework with data, a more comprehensive investigation on data availability and data requirements for agri-environmental indicator is underway (under the lead of EUROSTAT). The results will be presented at the Gothenburg summit.
A preliminary overview on available information has been published in 1999 by DG AGRI in co-operation with EUROSTAT and DG ENV under the title *"Agriculture, Environment, Rural Development: Facts and Figures"* (the full text can be found on the web-site of DG AGRI).
Comprehensive work on qualitative and quantitative information relating also to sustainable agriculture and rural development has been carried out by the European Environmental Agency. Important publications in this respect are the *"Environmental Assessment Report No. 2 – Environment in the European Union at the turn of the century"* and *"Environmental Signals 2000"*.

5 June 2001

Invited Paper Session

FISHERY STATISTICS: CURRENT AND FUTURE CHALLENGES

Chair: R. Grainger

Norms and Standards for Coordinated Fishery Statistics Gathering

David Cross[1]
Eurostat, European Commission, BP 1907 Luxembourg
email: david.cross@cec.eu.int

Abstract: This paper reviews the achievements of the Coordinating Working Party for Fishery Statistics in developing harmonised concepts, definitions and methods for the collection, compilation and dissemination of fishery statistics at the international level.

Keywords: norms, standards, harmonisation, fishery, statistics

1. Introduction

Fishery statistics was one of the first subjects for which the importance of international collaboration was recognised. This arose from the realisation that marine resources were not limitless and that fisheries would have to be managed. Furthermore it was realised that this management would have to be undertaken on an international basis because there were few if any stocks of fish for which any one country had exclusive "ownership" and that, for this management to be efficient, the managers would have to be using the same concepts and definitions.

2. The Coordinating Working Party (CWP)

An Expert Meeting on Fishery Statistics in the North Atlantic Area was organised by FAO[2] and co-sponsored by ICES and ICNAF in Edinburgh (Scotland) on 22-29 September 1959 and was attended by participants from 14 countries and 6 inter-governmental agencies. The Expert Group undertook reviews of the requirements of users of fishery statistics and of existing discrepancies in the standards used by FAO, ICES and ICNAF. More importantly, it initiated the joint task of these agencies by proposing and encouraging, for a large variety of sectors within the overall fields of national, regional and global fishery statistics, harmonisation in the following important sectors of work in fishery statistics:

- concepts and their definitions,
- classifications of craft, gear, species, areas, etc.
- methods (including harmonised questionnaires),
- simplification of reporting procedures.

---

1 Chairman, Coordinating Working Party on Fishery Statistics.
2 See annex 1 for the list of acronyms.

The Expert Meeting concluded its report (Anon., 1962) by inviting "the Governments of Canada, the Federal Republic of Germany, Iceland and the United Kingdom; and ICES, ICNAF and FAO to appoint one expert to form a continuing Working Party on Fishery Statistics in the North Atlantic Area with the power to co-opt additional members to help them in their task."

The Tenth Session of the FAO Conference (Rome, 31 October - 20 November 1959) approved this recommendation. Subsequently ICES and ICNAF agreed to participate and support the CWP, as the Continuing Working Party on Fishery Statistics for the North Atlantic Area came to be known.

Since that time the CWP has performed its appointed tasks and has increased its membership to include other international organisations with remits and responsibilities beyond the North Atlantic. The current members of the CWP are CCAMLR, Eurostat, FAO, IATTC, ICCAT, ICES, IOTC, IWC, NAFO, NASCO, OECD and SPC[3]. While retaining the acronym "CWP" its name has changed successively to the Coordinating Working Party on Atlantic Fishery Statistics in 1969 and then in 1992 to the Coordinating Working Party on Fishery Statistics.

Currently the CWP holds formal sessions at 2-3 year intervals, these sessions being attended by the representatives of the secretariats of its member agencies and of the statistical reporting offices of member countries of these agencies. This national representation is co-ordinated by the agencies to give as broad as possible a spectrum of national statistical systems and experiences.

A more complete account of the CWP, its origin, role and structure may be found in FAO Fisheries Circular no 903 (FAO, 1995).

3. Concepts and definitions.

In its early days particularly, the CWP paid a great deal of attention to the development of harmonised concepts and definitions. It is important that scientists, statisticians, administrators and other fishery workers in international fora should be employing the same terminology. Early CWP discussions made great advances in the definition and application of such concepts as the nominal catch (the live weight equivalent of the landings) and the attribution of nationality to catches, indeed so much so that these concepts and definitions have been applied universally.

Even now, over 40 years later, the CWP regularly reviews these concepts and definitions to ensure that they are still applicable and, where necessary, modify them to take account of changes in fisheries. An outstanding example of this is with aquaculture, an activity that started out as little more than providing fish with a sheltered environment and additional food and has developed in recent years into a highly specialised and technical industry. The definition of aquaculture has been reviewed and modified several times to take account of these changes in the industry.

[3] The International Commission for the Southeast Atlantic Fisheries (ICSEAF) was a CWP Member Agency until it ceased its activities in 1990

4. Classifications of craft, gear, species, areas, etc.

The CWP has developed classifications covering a number of topics. Examples of these classifications are:

- International Standard Statistical Classification on Aquatic Animals and Plants (ISSCAAP),
- International Standard Classification of Fishing Vessels (ISSCFV),
- International Standard Classification of Fishing Gears (ISSCFG)

These classifications are often used outside of the statistical domain by fishery managers and administrators. For example, the ISSCAAP includes a three-alpha code for fish species that is used widely, both nationally and internationally, in the reporting of fish catches on fishing log-books and other administrative documents.

5. Methods (including harmonised questionnaires)

One of the greatest achievements of the CWP is in the work of harmonising of questionnaires used in the transmission of fisheries data by national authorities to international agencies. The national authorities frequently have to send the same data to more than one agency. For example, an EU Member State fishing in the Northeast Atlantic has to send catch statistics to ICES, Eurostat and FAO. The CWP has developed a series of paper questionnaires (the STATLANT questionnaires) that result in the national authorities sending the same completed questionnaire to each of the concerned international agencies. While there is a separate questionnaire for most of the FAO major fishing areas, the questionnaires have a standard presentation and use the same concepts and definitions: the major differences are the species for which catches are recorded. The STATLANT questionnaires greatly reduce the workload on the national authorities and ensure that each organisation receives the same data.

In the context of its membership of the CWP, Eurostat's programme of fishery statistics has translated these STATLANT questionnaires into EU legislation and the other CWP members have accepted that the electronic format used by EU Member States to submit these data to Eurostat meets their requirements.

Despite the common system of questionnaires, the data-bases of the CWP agencies frequently contain differing data: a common cause of this is the failure of the national authorities to send revised or up-dated data to all the concerned international agencies. The CWP has noted the importance of eliminating discrepancies between their respective data-bases and regular exchanges of data between the agencies has become a standard procedure.

A very recent example of the collaboration between the CWP agencies has been the development of an integrated data-file for the catch statistics of the Atlantic Ocean. This file, to be made available on the FAO web-site and on CD-ROM and consultable using FAO's FISHSTAT Plus software, brings together in a common format the Atlantic catch statistics held on the data-bases of CCAMLR, CECAF, FAO, GFCM, ICES and NAFO for the period 1950-98.

6. Simplification of reporting procedures.

While the development of the harmonised STATLANT questionnaires has been a great achievement of the CWP, the hardcopy (i.e. on paper) reporting of data is an inconvenience for most national authorities requiring the extraction of data from computer files. The CWP has developed common standards for the submission of data in computer files and actively encourages the national authorities to use these formats for the submission of data and, more generally, for the exchange of data.

7. Is the CWP's work finished?

Certainly not! From remarks in the above sections it will be seen that the CWP has a continuing role of monitoring the definitions, concepts and working methods that it has established and coordinated statistical programmes amongst agencies. However, this is not the only work to be undertaken.

Until fairly recently the collection and compilation of fishery statistics was heavily biased towards the data for management of the biological resources. With the general acceptance that there is an over-capacity in the fishing fleets, a better integrated approach to fisheries management has to be developed, taking into account the biological, environmental, social and economic aspects of fisheries. The CWP will be required to develop, with the assistance of qualified experts, the concepts and definitions for the parameters to assess these additional aspects of fisheries and to determine how these additional elements can be assimilated in the existing statistical programmes.

The lack of adequate and reliable capture fishery and aquaculture statistics for many countries, and for developing countries in particular, is an extremely serious problem which the CWP may have to address. The CWP could provide guidance on the establishment of cost effective and sustainable data collection schemes and methodologies and maybe could also be influential in motivating donor support for statistical development projects and training.

References:

FAO (1962) Report of the Expert Meeting on Fishery Statistics in the North Atlantic Area. *FAO Fisheries Report* no 3. Rome, FAO. 26 p.

FAO (1995) The Coordinating Working Party on Fishery Statistics: its origin, role and structure. *FAO Fisheries Circular* no 903, Rome, FAO. 14 p.

Annex 1: Acronyms for international agencies

CCAMLR	Commission for the Conservation of Antarctic Marine Living Resources
CECAF	Commission for the Eastern Central Atlantic Fisheries
Eurostat	Statistical Office of the European Communities (European Commission)
FAO	Food and Agriculture organisation of the United Nations
GFCM	General Fisheries Commission for the Mediterranean
IATTC	Inter-American Tropical Tuna Commission
ICCAT	International Commission for the Conservation of Atlantic Tunas
ICES	International Council for the Exploration of the Sea
ICNAF	International Commission for the Northwest Atlantic Fisheries (predecessor of NAFO)
ICSEAF	International Commission for the Southeast Atlantic Fisheries *(ceased its activities in 1990)*
IOTC	Indian Ocean Tuna Commission
IWC	International Whaling Commission
NAFO	Northwest Atlantic Fisheries Organisation (successor to the International Commission for the Northwest Atlantic Fisheries - ICNAF)
NASCO	North Atlantic Salmon Conservation Organisation
OECD	Organisation for Economic Cooperation and Development
SPC	Secretariat of the Pacific Communities

Sustaining Fishery Statistics Programmes in Developing Countries

Romeo S. Recide
Director, Bureau of Agricultural Statistics,
Department of Agriculture
BEN-LOR Bldg., 1184 Quezon Avenue, Quezon City, Philippines
e-mail: rsrecide@mozcom.com

Abstract: Despite the urgent and serious demands for high-quality statistics on fisheries, efforts to develop the statistical systems to produce these statistics have not made significant inroads. Experience in the region has shown that the main obstacles to sustainable fishery statistics development are inadequate funding, insufficiency of qualified manpower and poor facilities for data processing and dissemination. Among the measures recommended to support the sustainability of national fishery statistics programmes are campaigns to advocate more funds for these programmes, formulation of realistic development plans, development of cost-effective and sustainable data collection schemes and methodologies, and human resource capability building.

Keywords: Fishery statistical programmes; funding, manpower and facility constraints; fishery statistical frameworks; advocacy campaigns; long-term development plans; sustainable methodologies; manpower development.

1. Introduction

Since the early 1990's, a number of fora in the Asia and Pacific Region have repeatedly pointed-out the pressing need to insure the availability and accessibility of reliable and timely statistics for formulation of fishery policies and management plans. The need has not diminished over the years. Adequate data are, in fact, becoming even more critical for fishery sector management due to increasing competition for access to resource bases, intensified exploitation and declining abundance of natural fish stock[1].

The urgency and seriousness of the call for availability of high quality statistics notwithstanding, it is noted that strides in the fishery statistics improvement process of many countries have actually not been successful. Up until this time, many national statistical programmes, especially in developing economies, have remained relatively weak. In particular, they are reportedly still reeling from serious concerns such as the prevalence of weak and inconsistent data sets, significant gaps and the issue of relevance of existing data series to users' demands. Among the reasons commonly cited for these noted weaknesses are the lack of skilled personnel to collect and compile statistics, poor systems of data collection, validation and maintenance, inadequate resources and non-sustainability of development efforts.

[1/] FAO/SEAFDEC.1997. Regional Cooperation for Fishery Statistics Programmes. Paper presented at the FAO/SEAFDEC Regional Workshop on Fishery Statistics in Asia, held in Bangkok, Thailand on 19-23 August 1997.

This paper reviews some of the most commonly reported constraints that impede the development and maintenance of fishery statistics. It then outlines the strategic courses of action that can be adopted to promote a sustainable fishery statistics programme in developing countries.

2. Issues and Constraints in Fishery Statistics Development

As experienced in many countries, fishery statistics development is a slow and a seemingly never-ending process. The Philippines, for example, has a National Fishery Statistical System (NFISS) that has been in existence for over half a century. Yet its NFISS is still far from being categorized as a fully developed system capable of responding to all the data needs of users. Its inadequacies maybe partly attributed to the ever-increasing interest in fishery statistics which has consequently imposed more demands than what the system's resources can supply. Also, these (inadequacies) can be traced to the dynamic nature of the fishery sector wherein new concerns are continuously emerging and thus, the needs of data users are constantly burgeoning and changing. There are, however, other endogenous as well as exogenous factors that more seriously constrain the full development of a national fishery statistical programme. The three most significant factors are: (i) lack of funds, (ii) lack of qualified manpower, and, (iii) inadequate facilities for processing and dissemination of fishery statistics.

(i) Lack of Funds

Funding is the most serious area of concern being encountered by all developing countries in institutionalizing fishery statistical programmes. In these economies, statistical agencies usually get a very limited share in the allocation of government financial resources. Making matters worse is that within the statistical system itself, the statistical programmes for agriculture and fisheries are often given the lowest priority in the sharing of limited funds.

The inadequacy of funds allotted to the fishery statistical programme affects not only the frequency of statistical surveys. More significantly, it tends to affect the quality of the generated information. In most of the locally-funded statistical activities, costs become the over-riding constraint in the design of activities. Thus, while the standard theory of sampling would require an optimum number of samples to satisfy the requirements of efficiency, representativeness, reliability and flexibility, the adopted designs usually neglect the problems on bias, non-sampling errors and undercoverage of sampling units because of budgetary limitations. The information generated from these makeshift methods are likely to be far from being accurate and comprehensive.

Lack of funds also precludes the updating of sampling frames which further contributes to inadequacies in sampling strategies. An outdated sampling frame can lead to problems of coverage and underestimation which in turn will result in incomplete, inaccurate or unreliable data.

But perhaps the biggest setback engendered by fund inadequacy is the inability of statisticians in developing countries to engage in statistical methodological development activities. Without the opportunity to try out new and more efficient methods, they are consigned to repeated application of poor systems of data collection, validation and maintenance.

Even the implementation of foreign-assisted projects is not a guarantee that the fishery statistical programme could be sustained in a developing country. In many instances, it has been observed that statistical development efforts would virtually come to a standstill upon the termination of the project. The counter-part budget provided by the national government is, more often than not, insufficient to sustain the implementation of all the statistical activities that have been developed during the existence of the foreign-assisted projects. Without the funding needed, the national fishery statistical programme would gradually revert to the traditional methods of doing things.

(ii) Lack of Qualified Statistical Manpower

In many developing countries, the scarcity of personnel qualified to handle the responsibilities related to the generation of fishery statistical programme is manifested in the absence of a sound statistical framework which can be used to guide the formulation of the fishery statistics programme and to assess its adequacy and relevance. The absence of a good statistical framework, in turn, may lead to the generation of data sets that are irrelevant to the needs of users, on one hand or the non-generation of required data, on the other. Also, the lack of framework leads to the slow development of the system as it fosters discontinuity of ideas and incoherence in approaches. Hence, neither improvements nor innovations could be incorporated into the system over long periods of time.
Many statisticians managing the fishery statistics units are often deficient in knowledge in fishery production and resource economics. Their low level of understanding of the technical aspect of fishery can be a serious impediment in the conceptualization and development of improved statistical methods and procedures.

(iii) Inadequate Facilities for Processing and Dissemination of Fishery Statistics

Another weakness of fishery statistical programme in developing countries is the untimely release of data. This is mainly attributed to the inadequate facilities that are made available for electronic data processing and transmission of fishery survey results. The lack of appropriate facilities adversely affect data quality when transcription and transmission errors creep into the data processing stage. On the other hand, inadequate information and communication technology facilities seriously diminish the timeliness and ultimately, the usability of fishery statistics.

3. Recommended Strategic Actions for Sustaining National Fishery Statistics Programmes

The improvement of fishery statistical programme in developing countries is now a matter of utmost necessity. Considering the significance of fishery in the lives of millions who are dependent on this sector, the various sectoral development programs and projects should be fully supported by adequate information to insure their successful implementation. The need for sustaining the fishery statistics improvement is further underscored by the growing competition within the sector as well as the necessity for enhancing the efficiency in its resource use. To move toward this direction, the following strategic actions are recommended:

3.1 Conduct of vigorous advocacy for bigger funding allocation to fishery statistical activities

The development of the national fishery statistical system of a country would continue to move along a roller coaster track unless adequate resources are appropriated to it on a continuing basis. A major step towards getting bigger budget for the system is to conduct a massive information campaign on the critical role of adequate fishery statistics in development planning and decision making in both the public and private sectors. The objectives of the various fishery-related statistical activities as well as the wide range of applications of their results should be clearly understood and appreciated by legislators and administrators of concerned fishery agencies.

An effective method of promoting appreciation of fishery statistics is to involve the private sector particularly the fisherfolk, traders and processing groups. Once they are convinced on the benefits to be derived from the outputs of different fishery surveys, they might even take the lead in conducting advocacy measures to increase funding for the statistical system. The partnership with the private sector can likewise insure that the generated fishery statistics will be in the form and standards as needed by users.

The assistance of international organizations like the FAO and the Asian Development Bank (ADB) may also be solicited in the advocacy campaign. These prestigious organizations are highly influential and positive results are usually achieved whenever they get involved.

3.2 Formulation of the medium and long-term development programme for fishery statistics

The national fishery statistical system needs a basic document that will provide strategic direction for its development in the medium and long terms. The programme must embody the priority statistical activities that can generate the required data sets and the planned improvements to enhance the capability of the system in responding to emerging data needs. It must also highlight the problems and issues confronting the system. Moreover, it should clearly spell-out its objectives, the strategies to be pursued to attain its set goals as well the estimated budgetary requirements.

This document is envisioned to serve not only as a development guide to statistical producers but it should also be useful as the basic information material for advocating bigger funding appropriation for the system.

3.3 Development of cost-effective and sustainable data collection schemes and methodologies

Due to the usual funding constraints, developing countries should not depend entirely on formal statistical surveys for generating fishery data. Rather, they must explore all possibilities for maximizing the use of administrative reporting system in collecting the desired information. Among the data sets that can be compiled through the administrative-based reports are the following:

3.3.1 FishingVessel Statistics

An amount of information can be compiled from the administrative forms used in the licensing and registration of fishing boats. The data items include the total number of fishing vessel by type, by tonnage class, gears used and number of operators. The list of fishing vessel statistics can also be easily expanded through the modification of the registration/licensing forms.

3.3.2 Commercial Fishery Statistics

As experienced in Malaysia, Thailand and Taiwan, fishery production data can be effectively generated through the Log Book System. In said countries, the Log Book is used for the detailed enumeration of the deep sea fishing vessels. The operators are required by regulation to fill-in the Log Book and submit it to the country's Department of Fisheries on a monthly basis. Aside from being a source of commercial fish catch data, the Log Book also serves as monitoring and assessment mechanism for the performance of deep sea fishing vessels[2/]. This scheme can well be used in other developing countries. The only requirement is a law or an administrative fiat that will strictly require commercial fishing vessels to regularly report their fish catch. Each fishing vessel shall provide to the concerned department the following reports duly certified by the vessel's Captain and transmitted monthly using the prescribed logsheet forms:

- A record of daily fish catch by fishing trip and fishing area, such fishing area defined by specific measurement of latitude and longitude positions; and,
- Daily record of quantity and value of fish catch by major species, spoilage, landing points, transshipment and/or other means of disposal.

The above-reports should be transmitted regularly to the officer or representative of the concerned monitoring agency at designated landing points.

3.3.3 Aquaculture Statistics

Aquaculture is now the source of new growth in the fishery sector of many developing countries. Due, however, to the wide range of activities and culture environment it covers, gathering of the requisite statistics for the purpose of periodically measuring the actual performance of this sub-sector poses serious financial and operational problems[3/]. Thus, a combination of surveys and administrative-based reports may prove necessary for the continuous monitoring of the sub-sector.

2/ Country Paper presented at FAO/SEAFDEC Regional Workshop on Fishery Statistics in Asia, Bangkok, Thailand, 19-23 August 1997.

3/ For purposes of economic accounting and reporting for instance, the various types of aquafarm whose operations must be periodically measured are classified as: (i) brackishwater fishpond, (ii) freshwater fishpond, (iii) freshwater fishpen, (iv) freshwater fishcage (v) marine fishpen (vi) marine fishcage (vii) oyster farm, (viii) mussel farm, and, (ix) seaweeds.

The above-cited scheme of gathering aquaculture data can soon be adopted in the Philippines due to the implementation of the Philippine Fisheries Code. This Code mandates the compulsory registration of fish hatcheries and private fishponds with the local government units where they are located. It also requires the country's Department of Agriculture (DA) to conduct a yearly inventory of all fishponds, fishpens and fishcages whether in public or private properties. Moreover, it stipulates that all fishpond, fishpen and fishcage operators shall annually report to the DA their production of fish and other aquatic products by species and types/method of culture and input utilization[4/] .
The above-described reporting system for aquaculture can easily be replicated in other countries.

3.3.4 Municipal Fishery Statistics

The collection of municipal fishery statistics through the administrative reporting system can be made possible by enacting and strictly enforcing ordinance on the compulsory registration of municipal fisherfolk and fishing vessels. The ordinance can require, among others, that the local government unit (LGU) should maintain a registry of municipal fisherfolk who are fishing or may desire to fish in municipal waters for the purpose of determining priorities among them, of limiting entry into the municipal water and of monitoring fishing activities as well as fish catch.
The LGUs shall also maintain a registry of municipal fishing vessels by type of gear and other boat particulars. Boat operators must be required to prepare and submit reports on their daily fish catch by major species.

3.3.5 Foreign Trade Data

Statistics on imports and exports of fish and other aquatic products which are important indicators on a country's self-sufficiency level for these commodities can easily be collected by the customs office at the major exit and entry points. Importers and exporters are all required to accomplish a declaration form for their traded products. Among the foreign trade data that can be extracted by the statistical agency from the declaration forms are:

a) Quantity of imports/exports of fish and fishery products, by type; and
b) Import/Export of fish and fishery products by country of destination/origin.

3.4 Strengthening the Manpower Complement of the Fishery Statistical Agency

The maintenance of adequate fishery statistics and the implementation of statistical development activities require a multi-disciplinary staff complement. In particular, the fishery statistical force should be composed of personnel with strong background and training in statistics, economics and fisheries. This requirement must be taken into consideration when recruiting additional manpower for the statistical unit.

[4/] Philippine Fisheries Code of 1998.

A core technical group must be created among the existing personnel complement. This select group should then be provided with advance training and, if possible, opportunities to undergo graduate studies in statistics. After the completion of their studies/training, the group can be expected to take the lead in conducting developmental activities and maintenance of improved systems for the national fishery statistical programmes.

Other existing personnel should be provided with specially designed courses in fishery statistical operations to upgrade their technical capability. The training modules must at least cover basic statistical concepts and procedures, production economics and data analysis.

4. Conclusion

Sustaining a fishery statistical programme in a developing country is certainly a daunting but nevertheless a manageable task. First of all, it requires a firm resolve and commitment from the national statistical agency to improve the data system for fishery. From this resolve and commitment should arise the strategies to transform the system from a weak state into a pro-active system capable of meeting the recurring and emerging information needs of its clientele. The critical mechanisms and infrastructures that can support the system's development in the long run should be immediately put in place. These include the preparation of long and medium term development plan and perspectives for fishery statistics, strengthening the technical capability of statistical personnel through training and selective recruitment and upgrading of processing and communication facilities. In addition, the statistical agency should also start, as soon as possible, the implementation of a more vigorous and systematic advocacy campaign for bigger funding. It must also intensify its present efforts of demonstrating to data users, particularly the business groups and legislators, its institutional capability to generate statistics even with funding constraints. This can be done through the development of cost-effective and sustainable data collection schemes and methodologies like the complementation of statistical surveys with the administrative reporting system for generating fishery statistics. Once it succeeds in convincing the target groups that it is a cost efficient data producer, campaigning for regular and bigger budgetary appropriations to fishery statistics should become less difficult.

Finally, international organizations can and should play a bigger role in the development and maintenance of fishery statistical programmes in developing countries. After all, fishery stocks and other resources often transcend national boundaries and their proper management and utilization require regional or international cooperation.

References

Congress of the Philippines. Philippine Fisheries Code of 1998 (R.A.8550). Manila.1998.

FAO/SEAFDEC. Regional Cooperation for Fishery Statistics Programmes. Paper presented at FAO/SEAFDEC Regional Workshop on Fishery Statistics in Asia. Bangkok, Thailand. 190-23 August 1997.

FAO/SEAFDEC. Various papers presented at the Regional Workshop on Fishery Statistics in Asia. Bangkok, Thailand. 19-23 August 1997.

Recide, Romeo S. Problems of Coverage and their Effects on the Accuracy of the Rice and Corn Survey. Proceedings of the Conference on the Problems of Generating Agribusiness Statistics. National Statistical Coordination Board. Makati, Metro Manila, Philippines. July 1993.

Recent Developments in the Philippines Agricultural Statistical System. Paper presented at the 16th Session of the Asia and Pacific Commission on Agricultural Statistics, October 28- November 1, 1996. Tokyo, Japan.

The Philippine Fishery Statistical System: Status, Constraints and Development Strategies. Bureau of Agricultural Statistics. Quezon City, Philippines.1998.

Monitoring Aquaculture Development: The Philippines' Experience. Paper presented at the Ad-Hoc Expert Consultation on Variables and Terminology for Aquaculture Monitoring in Asia. Bangkok, Thailand. September 1999.

Ramos, Candido M. Fisheries Statistical System in the Philippines. Department of Agriculture - Regional Field Unit I. San Fernando, La Union, Philippines.1988.

Ramos, Erlinda M. Status of Philippine Fishery Statistics Data System. Bureau of Agricultural Statistics. Quezon City. Philippines.1992.

Sustainable Statistical Development for Fisheries[1] Role of the International Institutions

Adele Crispoldi
Fishery Information, Data and Statistics Unit (FIDI)
Food and Agriculture Organization of the United Nations (FAO)
Via delle Terme di Caracalla, 00153 Roma, Italy
e-mail: Adele.Crispoldi@fao.org

Abstract: The paper reviews fishery statistical requirements and the special data needs for fisheries. It analyzes data collection aspects in the light of increased and emerging data demands for sustainable resources management. It discusses aspects of fishery statistical information systems that deserve renewed attention at international level, in particular aquaculture statistics. It summarizes the role of international institutions, like FAO, in promoting the adoption of standard international criteria that ensure the comparability needed for world fishery and aquaculture monitoring and sustainability[2] of data collection programmes.

Keywords: Fishery statistics, Aquaculture statistics, Data requirements, Data collection methods, Sustainable fishery statistical systems

1. Introduction

In recent years the demand for data and information concerning marine and inland fisheries and aquaculture has greatly increased. The increase is due to expanded demand for monitoring the implementation of sound fisheries policies, to ensure sustainable development of the resources, bio-diversity conservation and attention to environmental management, while maintaining the traditional contribution of fisheries to national social and economic goals. There are indications that the availability of natural fisheries resources for food and other uses has remained static at best, while demand for fish has continued to increase as a result of population growth and economic development. As aquaculture has been developing rapidly and is widely recognized as a sub-sector important for food security, national governments are encouraged to develop their aquaculture monitoring systems. There is a need for comprehensive and stable solutions of a sustainable nature to ensure that fishery statistical systems provide the basis for decisions that ensure that enduring benefits are derived from the use of the resources.

---

[1] The paper is sponsored by the Food and Agriculture Organization of the United Nations (FAO)
The concept of sustainability at national level applied to fishery statistics means the capacity of the system to achieve long lasting satisfaction of the sector's information need, and its ability to adapt and satisfy changing needs, while being economically viable.

[2] The concept of sustainability at national level applied to fishery statistics means the capacity of the system to achieve long lasting satisfaction of the sector's information need, and its ability to adapt and satisfy changing needs, while being economically viable.

International institutions have been promoting the standardization of concepts and definitions, the use of common classifications and the collection of data of known reliability, to increase international comparability of fishery statistics at all levels, and ensure the collection of data of global relevance. The information required for fishery and aquaculture development and management has been widely discussed at international level and conceptualized by national and international institutions. The methodology for collecting, processing and analyzing fishery data have also been conceptualized and the use of sound statistical methods for data collection has been internationally promoted; the improvement of the methodological basis for fishery data collection in developing countries has been supported -directly and indirectly- through national and regional projects and technical advice.

2. Need for fishery statistics

Statistics are essential for well-informed decision making at all levels and the collection of comprehensive and reliable fishery statistics is a pre-requisite for policy decisions taken by the government or by the fishing industry itself. Many fish resources, whether marine or freshwater, are highly migratory or straddle boundaries of national jurisdiction and the high seas. International definitions and classifications are needed to make sure that data collected from different countries are consistent, so that it is possible to make comparisons between statistical series.

In the United Nations International Classification of all Industrial Activities (ISIC) [3], which provides a framework for classifying all economic activities and facilitates international comparison of national statistics, fishery activities are in Class 0500 of Division 05 "Fishing, Operation of Fish Hatcheries and Fish Farms; Service Activities Incidental to Fishing**".** In that standard economic classification, fishery is defined as a productive activity encompassing the harvesting for consumption of living aquatic resources from natural stocks and from fish farming activities. The definition emphasizes the output of the industry.

The fishing industry is based on the exploitation of a natural resource not controlled directly by man except in fish farming and other aquaculture practices. Because of the lack of direct control and because the sea is a foreign environment, there is a relatively greater need to study and monitor its changes and to collect routinely a wider array of data than it exists for land based industries. With natural fish and shellfish resources stagnating or declining in many water bodies and seas and oceans, and consequently stagnating or declining fishery production -due to excessive pressure and exploitation, but also due to exogenous factors (e.g. coastal pollution, climatic influences, alteration of the natural ecological balance, conflicts over the use of coastal resources)- the industry is increasingly turning to the farming of the waters. Many operators have turned to aquaculture practices both to enhance the natural resource base of a given fishery (by way of e.g. re-stocking of water bodies and management) and to cultivate aquatic organisms, as farmers do for agricultural and livestock products.

In industrial and semi-industrial fisheries, statistical data are relatively easier to collect since submission of data by fishing companies is usually part of licensing. Most of the difficulties in data collection concern small-scale fisheries, which consist of large

[3] UN (1990) International Standard Industrial Classification of all Economic Activities (Series M, No 4, Rev 3) New York

numbers of small and dispersed units –some without fishing craft- operating on subsistence or semi-commercial level. Small-scale fisheries, inland and marine, constitute a most important sector in the majority of developing countries and this implies that there is a need for promoting the use of well-defined, cost-effective and sustainable sampling methods and techniques for collecting basic fishery data.
In order to evaluate the performance of fisheries in relation to management and development objectives at national level, and to fulfil regional and international requirements, statistical data of known reliability are required. In particular a sound national statistical programme for the sector would fulfil:

a) the need of ensuring that appropriate indicators can be formulated from raw data for monitoring the sustainability of fishery development;
b) the need for data of a wider scope, not exclusively of biological nature, to take into account social and economic factors influencing fisheries;
c) the responsibility to maintain control over the environment and bio-diversity for the conservation of fishery resources;
d) the need of separating fisheries production from various fishery sectors (e.g. marine capture, inland waters capture, fish farming and other aquaculture practices; commercial versus subsistence and recreational activities, etc.) for better monitoring trends and requirements in fish supply.

3. The components of a fishery information system

A wide variety of data is required for resource management and for the managing of the industry. Each discipline which studies the fish resource and the industry requires different types of data. For practical purposes of data collection in fisheries there are six major *components* in a system on which data and information are required for management purposes:

1. Natural (environmental and physical): [seasonal and climatic conditions, rainfall and temperatures, and all other factors that affect the fish population, its reproduction and migration behaviour; length of coastline, surface of continental shelf, depths, currents, fishing grounds, river estuaries and coastal lagoons, rivers, lakes, and other water bodies, species occurring, and density etc]
2. Biological [growth rate, size, age, mortality, inter species relationships]
3. Technological [vessels, gears, equipment, fishing power selectivity etc.]
4. Economic [prices, consumption, output (=catch and aquaculture production), cost of harvesting and post-harvesting, fisher's income and income distribution]
5. Socio-cultural [numbers and social status of fishers and aquaculture workers, occupational mobility]
6. Institutional and legal

The priority attached nationally to fishery data collection to satisfy the information requirements of each of the above components will largely depend on perceived needs and resources. Most countries have some form of collection of basic data on their fisheries but the array of data collected, the methodological approaches and the way in which the responsibility for collection and analysis is organized varies greatly. Because

of the limitations in the amount of resources available for setting up and maintaining a statistical collection system, there must be a system of priorities for the different categories of data to be collected, and the data quality of precision, accuracy and timeliness must also be considered.
There is however a minimum set of data that are internationally regarded as essential for national fishery management and international comparability purposes, and whose periodical collection should be ensured, that is:

(a) Estimates of total production of fish, by species of major commercial importance, by area of capture, their utilization for food and feed, combined with data on imports and exports. These elements combined constitute the basis for estimating average national supply in total and per caput;
(b) Estimates of total value of fish production by species; it is essential in assessing the relative importance of the fishing industry within the national economy, and combined with costs provides an indication of income from fishing;
(c) Prices at landing (average producers price); this information, combined with data on operational costs can provide indices of fleet performance, while time series and trends are used in economic analysis and market studies;
(d) Catch and fishing effort by fishing grounds and species constitute the basic elements in the formulation of indices of abundance; time series of fishing effort are indicative of declining or increasing trends of fisheries performance in districts and regions;
(e) Catch rates by boat and gear categories; combined with data on size at capture, permit a large number of analyses relating to gear selectivity and indices of exploitation;
(f) The number and type of fishing vessels, and their size and capacity, is a measure of the level of effort put into the fishery; combined with catches and hours fished provides an indication of the performance;
(g) The number of people involved in fishing is also an indication of the effort and the cost of labour in combination with catch is an indication of performance.

4. Fishery data collection

The *sources* of fishery data are primary -or fishery dependent- or secondary, including fishery independent data. Primary sources are the fishers themselves, and the market. For large-scale vessels, logbooks are designed to collect data on the time and place of fishing, the effort expended, the catch by species. Information from logbooks is compulsory in many fisheries. Catch sampling programs are an important source of information; fish can be measured at sea by observers or at landing sites, by port agents; samples can be obtained to determine the species composition, the sex ratio, and the age composition of the catch. Observers can participate in trips of commercial fishing vessels to collect information that is often not reported in logbooks (e.g. on by-catches and discards). The sources of post harvest data are all persons involved in transactions, at landing sites, fish auctions and other secondary markets, at processing units, and consumers markets. Fishery-independent data are for example scientific surveys (e.g. carried out to determine fish abundance) and foreign exports by commodities.

The *methods* of data collection from primary and secondary sources relies on several methods which can be used in combination with each other or individually. The most frequently used are surveys (census and samples), observations (site visits, oral interviews to collect specific, systematic and quantitative information), experiments (stock assessment, species abundance in time and space, growth characteristics), key informants (important in the case of small scale fisheries where official data are rarely collected in a systematic way).
The instruments of data collection are questionnaires (structured, unstructured), personal interviews, maps, data sheets, tapes, photographs and satellite imagery, GIS, GPS for vessels, administrative records etc. Similarly to agriculture *statistical procedures* are census, sampling [e.g. based on inventories of fishing units or fishing establishments]; sampling techniques include probability, random, systematic, cluster, stratified, non-probability samples; *data characteristics* include timeliness, level of detail, accuracy, coverage and completeness, internal consistency, accessibility to users, credibility of the data collection process and of the management process that uses the data.
The information required for fishery and aquaculture development and management has been widely discussed and conceptualized by international and national organizations; the methodology for collecting, processing and analyzing fishery data have also been conceptualized and the use of sound statistical methods for data collection has been internationally promoted.

5. Problems and constraints and the role of FAO in resolving them

The importance of reliable statistical fishery information and data in management and planning as well as in assessing the contribution of fisheries to food security, though internationally recognized, has not been addressed adequately in several countries. This is due to a large extent to the lack of proper human and financial resources required for conducting surveys and studies on fishery resources and their utilization.
It is desirable that all inputs (e.g. commercial fisheries, subsistence and recreational catches, surveys of catch per unit of effort) should be measured nationally with the same precision, rather than having a measurement of great or known precision, and all others measured approximately. Often national statistical programmes view the optimization of sampling needs in terms of the wider statistical programme. For composite fisheries there is the need of good co-operation in data collection among e.g. commercial fishers, subsistence and recreational fishers, regional and state agencies, so that the quality of the data is improved and the credibility of the management decisions taken on its basis and the policies established also increases.
Since the process and promotion of the use of the data is costly, time consuming and requires trained personnel, a good organization, and adequate statistical facilities and infrastructures, there is the need for a balanced and co-ordinated effort at national level so that through dynamic interactions between users and producers of the information the system meets the changing data demands effectively and efficiently.
To help meet increased and new data requirements needs, FAO has been trying to assist countries in upgrading their data collection, processing and reporting capabilities and to help guide, co-ordinate and monitor all of the many activities categorized as "fisheries" to improve their system. Guidelines have been published, technical advice and technical support have been provided bilaterally and also by means of technical consultations,

reports, training courses and workshops. Their role as often has been catalytic in improving national fishery statistical monitoring systems. Recently in the Asia/Pacific region, whose fisheries combined account for one third of the world total production, several workshops have been held to discuss possible improvements to the national data systems that could also yield better regional and international fishery and aquaculture data.[4]

Some of the *Concepts* which international institutions have promoted and which have a bearing in final estimates of world fishery production and trade are:

nominal catches versus landings
nationality of catches
statistical treatment of the discards at sea
statistical treatment of direct foreign landings
statistical treatment of subsistence and recreational fisheries
definition of aquaculture and separation of farmed production from wild production

6. The Importance of Aquaculture Statistics and Monitoring

Although growing of some aquatic organisms has been a traditional activity practised for several centuries in some countries, aquaculture is a fairly new industry in many other countries. The culture practice is becoming world-wide both in coastal and inland environments, and the culture of aquatic organisms is evolving rapidly through the diversification of species being reared under controlled aquatic environments.

At the international level, the lack of a robust set of structural information, production and other economic data related to the aquaculture sector as separated from fisheries was notably noticed in 1984 when the World Conference on Fisheries Management and Development was held. Such lack had gradually become a constraint to the analysis of global trends for the fishery sector as a whole.

The monitoring of aquaculture is very important for aquaculture management and planning, and requires a good statistical basis. As such, at national level statistical work should not be isolated, but rather, co-ordinated with aquaculture development, management and planning at both central and field levels. Priority should be given to strengthen national systems as they basically serve the national needs. Based on good model national system, an internationally comparable data collecting system would gradually be developed.

Fisheries and aquaculture may be viewed as a continuum of increasing levels of human interventions starting at one end with capture fishing (the simple extraction of a wild stock from its natural environment) and at the other extreme, fish rearing in completely man-controlled environment. Dividing neatly the two activities is not simple and has elements of subjectivity, but over the years, through dialogue and consultation with other international fishery bodies and individual experts, a definition of aquaculture for

---

[4] -*Ad hoc* Working group of experts in capture fishery data collection, Asia-Pacific Fishery Commission (APFIC), Bangkok, Thailand, 7-9 September 1999
-SEAFDEC/FAO Ad-Hoc Expert Consultation on Variables and Terminology for Aquaculture Monitoring in Asia, 13-16 September 1999, Bangkok, Thailand
-NSO Thailand/FAO - Workshop on Census of Agriculture 2000: Structural Aquaculture Statistics, 28 Feb.-3 March 2000, Songkhla, Thailand

statistical purposes has been elaborated and promoted by FAO. Although at the national level it is clearly the requirements of the users which will determine what will be included in the definition of the sector and what should be excluded, some internationally agreed basic definitions and classifications are needed to be used in different circumstances. The current FAO definition is the result of consultations that reflect the regional and national concerns that the harmonized definition should include three essential concepts, namely, the aquatic organisms reared and their farming environment, the aquaculture practice, and the ownership of the reared organism. In this regard, efforts have been made to promote nationally the use of the international definition of aquaculture.

The implementation of the recently approved identification of aquaculture, as separate from fisheries, in ISIC[5] –which formerly did not distinguish production units by *how* activities are carried out but rather by *what* is produced – will ensure that aquaculture statistics will be consistent with other economic data and lay the foundation for a harmonized treatment of aquaculture in national systems.

There is also the need for the harmonisation of concepts, terms and variables used in aquaculture to make the data and other information submitted to international and regional bodies comparable and reliable.

In several developing economies, the expansion of the aquaculture sector is closely related to agricultural activities and rural development. For example, reservoirs that are primarily used for irrigation are also used for aquaculture and waste products from agriculture are used as nutritional inputs into aquaculture. In rural holdings where aquaculture seems to be independent of agriculture, there can be a significant interchange of resources. Given the close relationship between agriculture and aquaculture, it is opportune to consider a closer link between agriculture and aquaculture statistics.

Work to further harmonize aquaculture terminology remains to be done and the implementation in data collection systems of fishery activities designated into particular practices, i.e., capture, enhanced, culture offer a useful guide and should be expanded to ensure harmonization, and to avoid double counting or missing out. A glossary of terms may be compiled and compared at the international level by FAO for dissemination to user groups.

7. National Actions for the Improvement of Aquaculture Monitoring

National governments are encouraged to develop their aquaculture monitoring systems. As human and financial resources for developing aquaculture monitoring are limiting factors, countries are encouraged to maximize the use of existing available data for the needs of multiple users.

To enable effective and efficient aquaculture management and planning, well-presented statistical and non-statistical information with analytical explanations should be provided as relevant and reliable time-series information. In order to produce such information, relevant institutions and experts should be involved in the analytical process. Countries may consider establishing a national multi-disciplinary co-ordination

[5] In March 2001 the United Nations Statistical Commission endorsed the proposal of the Technical Subgroup of the Expert Group on International and Social Classifications of splitting existing class 0500 into two new classes 0501 (Fishing) and 0502 (Fish farming).

mechanism to continuously develop and monitor aquaculture statistics programs at national and local levels and to ensure consideration of the scope of the data collected in view of changing data needs, both for production and planning.
Human resource development of the statistical personnel at different levels, particularly training of primary data collectors, should be encouraged.
Future requirements include data exchange systems that permit the electronic transfer of data to and from countries and FAO. The use of harmonized definitions is a pre-requisite. Countries that do not have a glossary at present are encouraged to develop one.
As a minimum the national fishery statistical system should be able to provide the following data for aquaculture:

a) Estimate of total production of fish, by species of major commercial importance, by aquatic environment (freshwaters, brackish waters, seawaters; coastal, off-shore).
b) Estimate of total value of aquaculture production by species; it is essential in assessing the relative importance of the sector within the national economy, and combined with costs provides an indication of income from aquaculture;
c) Prices at farm-gate level (producers price); this information, combined with data on costs can provide indices of productivity, and is used in economic analysis and market studies;
d) Production by species, by type of site (e.g. ponds, tanks, cages, pens, ropes reservoirs)
e) The number and type of installations, their location, size and capacity;
f) The number of aquaculture workers and labourers whether permanent or occasional.

8. Conclusions

Fishery problems have changed and the demands for statistics have also changed, with more emphasis being put on management and less on development, since most national fishery statistical systems were set up. Sound policy making and effective management require reliable fishery statistics. There is a need to collect systematically data on production from aquaculture practices. In the context of the principle of responsible fishing both within waters of national jurisdiction and on the high seas, new data requirements have been identified. Fishery statistics coverage should be improved and data quality should be checked through effective validation schemes. National fishery statistical services should evolve integrated statistical programmes in order to optimally utilize limited logistic and human resources, avoid duplication of activities and eliminate dissemination of conflicting statistics; in the case of rural aquaculture optimum use should be made of rural households surveys. The donor community should provide increased technical assistance in the field of fishery statistics to strengthen the national capability in using acceptable statistical methods for the collection, processing, storage, analysis and dissemination of fishery and aquaculture statistics.

References

K. Brander, (1975) Guidelines for the collection and compilation of fishery statistics, *FAO Fisheries Technical Papers N. 148*, 46 p.

FAO, (1985) Expert Consultation on the acquisition of socio-economic information in fisheries (with particular reference to small-scale fisheries). *FAO Fisheries Report N. 344*, 21 p.

FAO, (1990) Co-ordinating Working Party on Atlantic Fishery Statistics (CWP) Handbook of fishery statistics. Rome, FAO, Rome, pag. var.

Rana K.J., (1997) Guidelines on the collection of structural aquaculture statistics – Supplement to the programme for the World Census of Agriculture 2000, *FAO Statistical Development Series N. 5b*, 56p.

OECD, (1995) The significance of reliable statistics to conduct effective management – Workshop, OECD/GD (95) 55, Paris, 187p.

Eurostat, (1998) Fisheries- The collection and compilation of fish catch and landing statistics in member countries of the European Economic Area, *Methods and Nomenclatures,* Theme 5, 170p.

FAO, (1999) Guidelines for the routine collection of capture fishery data. *FAO Fisheries Technical Paper N. 382*, p. 113

SEAFDEC, (2000) SEAFDEC/FAO *Ad-Hoc* Expert Consultation on Variables and Terminology for Aquaculture Monitoring in Asia, 13-16 September 1999, Vol. 1: *Report of the Consultation*, Bangkok, Thailand, 36 p.

Ocean Studies Board/National Research Council (2000) Improving the collection, management, and use of marine fisheries data, National Academic Press, Washington D.C., 222 p.

Sustainable Statistical Development for Fisheries[1] Sample-Based Fishery Surveys

Constantine Stamatopoulos
Fishery Information, Data and Statistics Unit (FIDI)
Food and Agriculture Organization of the United Nations (FAO)
Constantine.Stamatopoulos@FAO.ORG

Abstract: The paper concerns data collection aspects in sample-based fishery surveys. It discusses circumstances under which sampling approaches provide cost-effective, efficient and sustaining solutions for the collection of basic fishery data, thus facilitating the development of statistics needed by fishery managers and planners. It also summarizes experience gained over the recent years in fishery statistical development undertaken by the Fishery Information, Data and Statistics Unit (FIDI) of the FAO. Major topics include constraints affecting the development and implementation of large-scale fishery surveys, methodological and operational tools that have been developed and implemented by FAO, and issues on sustainability and cost-effectiveness of national fishery statistical programmes.

Keywords: Fishery surveys, Data collection methods, Sampling aspects, Sustainable fishery statistical systems

1. Introduction

Among the major tasks of the FAO is the promotion of improved approaches and techniques for the collection of data on agricultural statistics, including fisheries and forestry. The need for reliable and comprehensive statistics has never been more acutely felt than at present, when most countries are engaged in large-scale planning. Large-scale statistical programmes, however, require a great deal of effort and funds for their development and implementation and these are major constraints for many countries with limited human and financial resources. The merit of the sampling approaches lies in providing a cost-effective and efficient method for the collection of data, thus accelerating the development of statistics urgently needed by fishery managers and planners.

The present paper summarizes experience gained over the recent years in fishery statistical development undertaken by the Fishery Information, Data and Statistics Unit (FIDI) of the FAO, with specific emphasis on sample-based fishery surveys.

Section 2 of the present paper provides background information related to the need for basic fishery data and the applications sectors where such statistics is of high importance and utility.

Section 3 deals with major constraints affecting the development and implementation of fishery statistical programmes and provides a number of indicators and criteria concerning national sustainability in maintaining medium and large-scale systems.

[1] The present paper is sponsored by the Food and Agriculture Organization of the United Nations (FAO)

Section 4 describes briefly FIDI's work in fishery statistical development over the recent years and discusses the relating methodological and operational tools that have been developed and implemented through its normative and field programme activities.
Section 5 outlines some major conclusions deriving from FIDI's experience in fishery statistical development and suggests points for discussion concerning sustainability and cost-effectiveness of national fishery statistical programmes.

2. Need for basic fishery statistics

Collection of basic data on catches, fishing effort, prices, values and other related information such as size at capture and length frequencies, constitute a key factor in a wide variety of applications. Sample-based fishery surveys that are conducted on a regular basis ought to be viewed not as an end in themselves but as an important source of fishery information of wide utility and scope.
It would seem appropriate to identify here a number of applications that depend directly on the availability of basic data resulting from fishery surveys. The list is not exhaustive but it essentially involves:

(a) Estimated total production of fish, combined with data on imports and exports, constitutes the basis for calculating per capita consumption of fish, which is subsequently used in the formulation of food balance sheets;
(b) Estimated total value of fish production is an important element in assessing the relative importance of the fishing industry within the national economy;
(c) Prices at landing, combined with data on operational costs can provide indices of fleet performance;
(d) Catch and fishing effort constitute the basic elements in the formulation of indices of abundance;
(e) Catch rates by boat and gear categories, often combined with data on size at capture, permit a large number of analyses relating to gear selectivity and indices of exploitation;
(f) Catch rates are often used for formulating indices of abundance for different fishing grounds;
(g) Time series of prices are used in socio-economic studies;
(h) Time series of fishing effort are indicative of declining or increasing trends of fisheries in districts and regions;
(i) Trends regarding the human involvement in the fishing industry can be formulated.

To help meet these data needs FAO has been assisting countries in upgrading their data collection, processing and reporting capabilities. Technical assistance at national and regional level is a significant component of the work programme of FAO's technical units responsible for fishery statistical development and involves both normative and field programme activities. Outputs of the normative activities include technical documents on statistical methodology and guidelines for data collection, while field programme activities involve project formulation and implementation, technical backstopping and organization of training courses and workshops.

The present paper focuses primarily on sample-based fishery surveys and underlines the need for well-defined and cost-effective sampling approaches with the view of assuring the sustainability of national fishery statistical programmes.

3. Constraints in data collection programmes

3.1 Human and financial resources

The backbone of a data collection programme is the team of field data recorders and supervisors. Well-trained and motivated data recorders is the principal concern of fishery statistical units since it is through them that primary information flows from fishing sites and markets up to the offices responsible for data processing and analysis. Consequently the quality and utility of the statistics produced is a direct function of the effectiveness and timeliness of field operations. Availability of sufficient and adequate human resources is often one of the major constraints in the implementation of medium and large-scale fishery surveys operated on a regular basis, particularly in cases of fishery administrations with limited budget allocated for data collection programmes.
Mobility of data recorders, that is their ability to visit as many locations as possible during their allocated time, is another aspect that affects the quantity of collected data as well as their representativeness. Lack of transportation means reduces statistical coverage and increases the risks of biased data, since data collection is always conducted at the same few locations. Mobility is also important for supervisory functions lack of which would leave data recorders on their own and without supervision and guidance.
Mobilization of human resources who are not necessarily regular staff of fishery administrations is often a good approach for obtaining information that does not require highly skilled personnel. For instance, data on the level of activity of fishing units (a basic parameter in estimating fishing effort), is sometimes obtained directly from the fishermen themselves; students or scouts resident in fishing sites can also act as data recorders thus increasing significantly the sample size at reasonably low cost.

3.2 Statistical and computing drawbacks

Experience shows that, in general, design and implementation of sample-based fishery surveys make little or no use of statistical indicators concerning sampling requirements that would guarantee an acceptable level of reliability for the estimated population parameters. On several occasions statistical developers tend to operate on an “over-sampling” basis, despite the fact that *a priori* guidance on sample size requirements is feasible in most cases. Lack of well-defined and cost-effective sampling schemes tends to increase the size and complexity of field operations without visible benefits in resulting accuracy and this, in turn, has a direct impact on the logistical aspects of data collection and data management procedures.
Excessive stratification of the target population without due consideration to its impact on sampling requirements can be another limiting factor. Much refined stratification certainly improves the homogeneity aspects of a population but also has serious impact on sampling effort and survey cost. This factor is at times overlooked by statistical developers who continue to apply old sampling schemes proportionally to the size of

newly created stratified populations and maintaining the same total number of samples collected over the reference period. According to basic sampling theory this approach is not appropriate and safe sample sizes ought to be reviewed and adjusted after stratification.

Lack of sufficient and appropriate data processing tools and methods are often a negative factor in the operations of fishery statistical programmes. Due to the wide spread of micro-computers and increased computer literacy among data producers and users, computer systems (of varying sophistication and robustness) have, in fact, become an inseparable component of fishery statistical systems but, with few exceptions, they tend to be fragmented, inflexible and heavily centralized. Lack of flexibility and robustness implies frequent interventions to the computer software thus increasing the chances of undetected programming faults, while lack of a modular system structure creates processing bottlenecks and deprives decentralized offices from the benefit of locally processing and analyzing their own data.

3.3 Sustainability of fishery surveys

Irrespective of the degree of complexity of the fishing operations monitored by a fishery statistical system, the essential elements of the data collection scheme and the population parameters involved in the estimating process are always inter-related by means of a general expression in which the estimated total catch is a function of an overall "Catch-Per-Unit-of-Effort" (CPUE) and the total fishing effort exerted. Estimation of total catch is the basis for the derivation of all other basic fishery data such as species composition and values of landings.

CPUEs and fishing effort are thus the two principal parameters and are usually estimated within the logical context of a reference period, a geographical area and a specific boat/gear category. Of the two parameters mentioned above, estimation of the CPUE constitutes an easier task for the following two reasons:

(a) Distribution of the CPUE is such that relating sampling approaches are more robust and less demanding than those for fishing effort;
(b) CPUE requires only one sample-survey, whereas fishing effort depends on two sample surveys and a census.

Although in most fishery surveys the estimation process is fairly simple from the mathematical standpoint, there are other aspects, both statistical and non-statistical, that necessitate the parallel existence of well-defined, robust, modular and flexible computer systems. Samples collected from the field would have to be classified and stored, estimation of population parameters ought to be as automated as possible to avoid lengthy, risky and routinely performed manual computations, automatic preparation of statistical indicators and diagnostics is essential in identifying problem areas and deciding as to the type of corrective action required.

It would thus seem appropriate to state that national fishery statistical programmes are sustainable and cost-effective when they make good use of available human and financial resources and estimate fishery population parameters with reasonable accuracy and precision. Sustainability depends also on the existence of robust and flexible data tools that facilitate the storage and processing of data, provide meaningful statistical diagnostics, and make available reporting and data dissemination functions at decentralized level. In this respect the aspect of statistical training at all levels (data producers and users) and that of computer training (data operators and analysts), is of primary importance since it guarantees the monitoring, co-ordination and evaluation of field and office operations, permits application of corrective actions if and when necessary and, above all, safeguards the continuity of ongoing statistical programmes irrespective of staff changes and turnover.

4. Technical assistance by FAO

4.1 Need for standardized statistical and computing approaches

Most data collection systems have many common characteristics, irrespective of their environment and individual methodological and operational aspects. Starting from this concept FIDI over the years has developed a family of standardized statistical approaches and computer software (Artfish), aiming at facilitating the design and implementation of shore-based fishery surveys on fish production and values.
Artfish stands for "Approaches, Rules and Techniques for Fisheries Statistical Monitoring". It was first introduced in 1994 as an early prototype running under MS-DOS and since then it has been implemented in about 15 countries, mainly in Africa and Asia. The principal advantage of using standard statistical procedures and software in medium/large-scale data collection systems is that a customized data collection system takes about three years to develop and implement, whereas development costs are drastically cut through the use of error-free and standard approaches and software.
In 1998 FIDI started the re-design of Artfish to run under Windows. Its final version was issued in June 2000 and consists of three components: Artplan for survey planning and training, Artbasic for handling primary catch/effort data and Artser for data integration and reporting. Artfish does not require additional programming, changing of software or any specific computer expertise. It has been designed to adapt to commonly used data collection programmes and its use can thus be as sophisticated as the country needs dictate. Users need only construct the required survey structures and feed the system with parameters and sample data.

4.2 Training aspects

Most of the training in survey design is conducted through the planning and training Artplan module. Artplan operates with empirical parameters and makes maximum use of existing knowledge regarding fishing operations and patterns. Its functions include:

(a) Setting-up working frames with sites, boat/gear types and strata;
(b) Definition of fishing patterns;
(c) Simulating fishing operations;
(d) Generating "pseudo-census" information;
(e) Evaluating alternative sampling scenarios for cost-effectiveness.

4.3 Survey implementation aspects

The Artbasic module of Artfish is used for the storage and processing of basic data on catches and values and fishing effort. It operates on standard classifications, frame survey data and samples on catches, fishing effort, prices and values. Its results consist of monthly estimates of fish production and values by species within the logical context of a calendar month (or a sub-period), a stratum and a specific boat/gear type. All Artbasic estimates are presented with the associated statistical diagnostics (such as variability explained in both space and time, expected accuracy level, etc.). Artbasic is usually operated on a decentralized basis thus offering research and reporting services nearer to the data sources.
Artser is the reporting component of Artfish. Estimates from Artbasic are automatically gathered and formatted under an integrated database structure that allows for flexible and user-friendly data screening and extraction, data grouping, reporting and plotting.

5. Conclusions and points for discussion

(a) There is currently a need for well-focused and unambiguous methodological and operational guidelines relating to the design and implementation of sample-based data collection programmes;
(b) Most of the related computer systems currently in use seem to have served their purpose well but their operating platforms should be changed so as to respond better and more sufficiently to current and future needs;
(c) Practical statistical training of statistical producers at all levels and well-focused computer training of data operators is of primary importance as these people constitute the backbone of regular data collection and processing systems;
(d) There is a need for well-focused technical assistance programmes at sub-regional and/or national level that can also give a good initial thrust to the development of fisheries statistics. This is stressed by the ongoing FAO/World Bank/USDA Collaborative Initiative for the improvement of agricultural statistics (including fisheries and forestry) that was launched in 1998;
(e) In recent years and through its Technical Co-operation Programme (TCP), FAO has financed a good number of small fishery statistical development projects in Africa and in Asia. Due to their size and duration these projects had limited scope and focused on implementing prototype systems. Consolidation and system expansion could well be the subject of larger projects if major donor countries and agencies would agree to financially support such activities.

References

Bazigos, G. P. (1974) The design of fisheries statistical surveys – inland waters. *FAO Fisheries Technical Paper*, 133.
Bazigos, G. P. (1983) Applied Fishery Statistics. *FAO Fisheries Technical Paper*, 135.
Caddy, J. F. and Bazigos, G. P. (1985) Practical guidelines for statistical monitoring of fisheries in manpower limited situations. *FAO Fisheries Technical Paper*, 257.
Cochran, W. G. (1973) Sampling Techniques. John Wiley & Sons, New York. 9-17.

Deming, W. E., (1960) Sample Design in Business Research. John Wiley & Sons, New York.
FAO, (1993) Report of the Working Group on Artisanal Fisheries Statistics for the Western Gulf of Guinea, Nigeria and Cameroon, 49.
FAO, (1999) Guidelines for the routine collection of capture fishery data. FAO Fisheries Technical Paper, 382.
Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953) Sample survey methods and theory. John Wiley & Sons, New York. p.58.
Stamatopoulos, C. (1996) Report of the Workshop on the use of a fisheries statistical software ARTFISH. FAO-IDAF Project Technical Report, 83.
Stamatopoulos, C. (1999) Observations on the geometrical properties of accuracy growth in sampling with finite populations. FAO Fisheries Technical Paper, 388.
Sukhatme, P.V. and Sukhatme B.V. (1970) Sampling Theory of Surveys with Applications. FAO, p.7-25.
Thompson, S. K., (1992) Sampling. John Wiley & Sons, New York. p. 11-34.

ESSAYS

n. 12/2002

Conference on Agricultural and Environmental Statistical Applications in Rome

Rome, 4-8 June 2001

Volume II

ISTAT

6 June 2001

Plenary Session II

QUALITY STANDARDS FOR AGRICULTURAL STATISTICS

Chair: T. Garcia Azcarate

Improving the Quality and Value of Statistics on Agricultural Holdings by Making Use of Administrative Data and Other Statistical Sources

Peter Helm
Ministry of Agriculture, Fisheries and Food, 1-2 Peasholme Green, York, YO1 7PX, United Kingdom, Tel.: +44 1904455240
e-mail: peter.helm@maff.gsi.gov.uk

Abstract: MAFF's e-business vision is to maximise the use of administrative data for statistical purposes and avoid duplicate collection. Using these sources can add "quality" to the statistics produced.

Keywords: E-business, June agricultural census, business registers, administrative sources (IACS, cattle tracing system, CTS), geographical information system, balance sheets.

1. Introduction

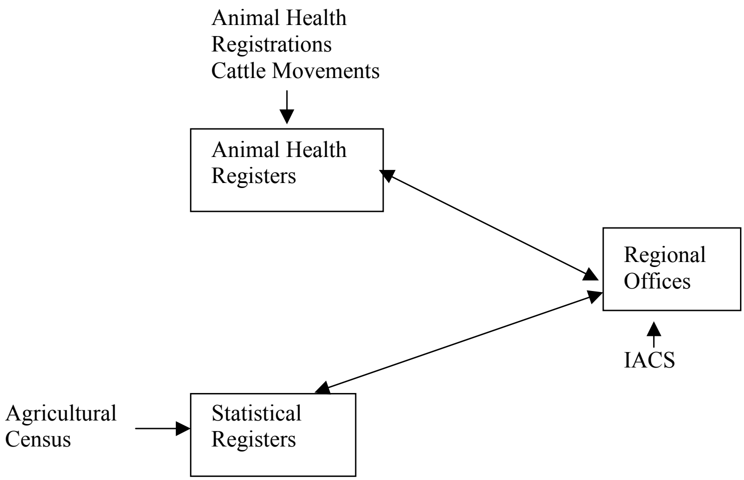
The aim of this paper is to explore the developing thinking and work on using administrative systems within the UK. While the initial push for this work has been to reduce the compliance costs on farmers, the interest in this paper is from the perspective of improving the quality and robustness of the figures.
Statistical surveys of farms and agricultural holdings are and will remain an important source of information on agricultural statistics. There are now increasing opportunities for improving these statistics by making greater use of administrative sources and checking the consistency with other statistical sources.
Hopefully the UK experiences will be of interest and value to statisticians in other countries. The work should be seen as being from the perspective of a country with a well-established statistical system. Retaining this as the base, we are trying to benefit from the alternative information sources available. It is well recognised that there is not a single model. Each country has its own context and systems.

2. UK Farm Statistical System

The collection of statistics on agricultural holdings centres on the annual agricultural census in June each year. Despite its name, a full census is only completed once every 10 years, with a large annual survey conducted in intervening years. Surveys are run along similar lines in each of the United Kingdom Agricultural Departments. MAFF has the responsibility for running the census in England and collating the UK figures. Information is collected on crops being grown, livestock numbers, labour employed and other basic details on the holdings. As well as providing statistics, the census is used to keep the farm registers up-to-date. It is used to produce well-targeted samples for more detailed surveys conducted at other times of the year.

The system of farm surveys has long been dependent on the administrative systems for the maintenance of farm registers. Within England, for example, responsibility for maintaining the holdings registrations rests with the MAFF regional offices. They issue and maintain registrations that are needed for grants and subsidies, animal health or statistical purposes. Inevitably the records are more difficult to keep up-to-date for agricultural holdings which have little administrative contact with MAFF, or where significant structural changes are taking place. Proactive register maintenance activity is being conducted by the statistical teams to fill in any gaps (e.g. poultry, pigs and horticulture) by approaching the large companies that own farms or contract out production to separate farms. MAFF is giving consideration to a more integrated register maintenance system.

Figure 1: *MAFF Holdings Registers*



As systems develop, there are increased opportunities for using the administrative data to improve the quality of the figures. This is in respect of identifying missing holdings on the registers and checking the quality of returns.

3. Use of Administrative Data

3.1 Advantages and Disadvantages

The main types of administrative data include:

- claims for grants and subsidies, e.g. arable area payments (e.g. IACS);
- animal health records, e.g. a database tracking animal movements.

Tax records are another possible source (but are not available to the UK Agricultural Departments). It is very clear in the UK context that, while administrative information can be used for statistical purposes, there are severe restrictions on how statistical data can be used for administrative purposes.
The main advantages commonly cited in using the administrative sources include:

- low additional cost for statistical purposes;
- returns are likely to be audited (and may be more accurate than the statistical reporting);
- may provide additional range of data (more frequently);
- welcomed by farmers (in reducing compliance costs);
- full coverage rather than a sample.

The main disadvantages include:

- coverage may not be complete, e.g. holdings not claiming;
- the claim may not reflect the full position;
- reporting units may be different, e.g. business level basis for IACS;
- definitions and timing may not be the same as the statistical ones (and may change over time);
- may not have all the statistical information that is required;
- may not be easy to get hold of the information, e.g. administrative systems not designed to provide this, complex, flow disrupted at short notice if emergency situation develops;
- not under the control of the statistician;
- priority in processing may be given to non-statistical uses.

In using administrative data within the UK, we have viewed it as important to maintain the statistical framework. That is, we are seeking to fill in any gaps, where we think it is appropriate, from the administrative sources. For example, this could include the identification of holdings previously not thought to be producing that commodity. Or it could include providing details of areas of crop from the IACS records where previous comparisons have shown that this is virtually equivalent to the statistical reports.
It can be more difficult using administrative data sources in aggregate, given the different definitions and the risk that definitions could change over time. While aggregate figures can be used for measuring overall trends in a single product over time, this does not meet the more detailed policy analysis for looking at how holdings are changing. Modelling approaches can be used for converting administrative data into the statistical equivalents.

3.2 Examples

3.2.1 Example 1: IACS

Additional information is being obtained from IACS to improve the range of data that can be provided by the statistical system.

Farmers submit annual IACS return by the middle of May each year to claim their arable area payments. This provides information by field on the crops being grown. The crop areas being covered on IACS and the June census are now very similar in total, although there are some variations by non-cereal crop types.
Initial work in using the IACS figures focussed on trying to reduce what was perceived as the duplicate collection of data. Previous exercises have identified 5,000 holdings where we have been able to stop the annual issue of forms because we can get virtually the same information from the IACS return. There are a further 6,000 holdings where we have been able to get all the land.
From improving the quality of statistical outputs, two major uses have been made of the IACS data: GIS system and business link.
As for GIS System a grid reference has been allocated to every agricultural holding on the census. This is allowing the census information to be presented more informatively in maps and to be more closely linked with other datasets. The grid reference collected in the IACS provided firm information for about 50% of the holdings. The remainder were approximated from the postcode of the address of the farmhouse or estimated from the parish location of the holding. It is no exaggeration to say that this has significantly improved the presentation of the data.
Business links represent another important use of the IACS data. There is considerable interest from the economists and others in having information at a business level. The IACS datasets have proved very helpful in this in allowing us to get fuller information on which agricultural holdings are linked together in the same business group. As resources permit, it is intended to update previous work so that we can produce information on our best approximation to a business level dataset for June 2000.

3.2.2 Example 2: Cattle Tracing System (CTS)

This system is helping us check the coverage of the statistical registers as well as providing additional information.
This is a new system, which has been set up across Great Britain (England, Wales and Scotland[1]) to track the movement of cattle. All farmers are legally required to register cattle they have on a holding basis and report any movements. A recent exercise has updated this for the older live animals, which were not originally in the system.
Arrangements have now been made to receive the CTS records on a quarterly basis, from July 2001, to be used for statistical purposes. Work being planned in relation to this includes:

- comparison at an aggregate level of cattle recorded on the statistical survey census and the CTS (on farm);
- investigation of a sample of major holding level differences;
- an attempt to devise a model which will allow data from the CTS to be used to estimate the full set of statistical data required. For example, the CTS does not distinguish between dairy and beef animals (although the breed is recorded) or “intentions” questions (e.g. female animals over one year intended for slaughter);
- examination of how various types of animals are changing and using this to develop an improved forecasting model.

[1] Northern Ireland has a separate system.

The use of the CTS data should help to assure the robustness of the statistical counts and improve the range of information for understanding the cattle sector. Working with the two datasets will have value for both sets. While the CTS may help to reduce the amount of information collected on cattle in the June and December agriculture surveys, it is difficult to see how this can be eliminated entirely given that the administrative systems do not collect all the required information. Also CTS will allow improved production forecasts to be produced, i.e. flow forecast rather than snapshot forecast.

3.2.3 Example 3: Pig Registers

The administrative systems have pointed us in the right direction for improving the register of holdings with pigs.

It has been recognised for some time that the statistical counts of pigs in June and December surveys have been understating the position. It is particularly difficult for the administrative and statistical systems to keep pace with the setting up of new pig enterprises, particularly the outdoor units. While it is difficult estimating the size of the undercounting, crude estimates can be made by comparing the slaughtering figures with estimates from the June and December figures. Various assumptions have had to be made. The comparisons suggest that the statistical count on sows and gilts in pig may be up to 10% (45,000) short in Great Britain.

Extensive work has been done by comparing the statistical registers with the administrative registers to detect any gaps. This is not easy given that there are doubts on the comparability of the reporting units between the two sources. This work has identified 6,000 pigs on units that were incorrectly being classed as temporary holding numbers. More importantly, the exercise discovered that the statistical registers are missing holdings from some of the big pig companies. This covers holdings owned by those companies or units that are contracted to rear pigs. We are working with the large companies to improve the coverage of our information. This is also having a knock-on effect of improving the administrative records.

4. Use of Statistical Data

Statistical data can be compared (benchmarked) with other statistical sources to check on its consistency. Agriculture is well placed for these checks due to the detailed statistics collected at various points in the production process and food chain. Where there are inconsistencies, judgements have to be made on which source is likely to be the best. While this analysis identifies where there might be a problem, it is not easy to determine what the cause might be. Some examples are indicated on how this has been used.

4.1 Example 1: Cereal Balance Sheets

Various statistical sources are used to create the balance sheet, trying to match the availability of cereals with the amount used. Any discrepancies in the sources become evident.

The assumption has been that the weakest component in the balance sheet is the estimate of amounts of straight grain fed on farm. This is an important component, accounting for about 7% of total wheat usage and about 33% of total barley usage. This estimate comes from the grain fed to livestock survey, which is widely recognised as a difficult survey to get right.

At the request of the customers, a lot of work has been put into trying to improve the estimates. Improvements have been made to the trigger questions in the June census to determine which holdings are feeding straight grain to their livestock. This has helped to target the survey more effectively. The questionnaire was re-designed to maximise response rate and the sampling was revised to improve coverage although more work is planned on this.

Currently the position is that if all the balance sheet errors occur in this estimate, it is underestimating the wheat figures by about 3% and the barley figures by about 10-20%.

With regard to further work, consultants have been employed to develop a model considering how much straight grain is produced, using the standard feeding patterns of animals. This will help to determine the viable range of the estimates and assess how many of the balance sheet errors might be due to this item. Consideration is being given to the design of the survey, considering the case for a survey that balances up production and usage at the farm level.

4.2 Example 2: Farm Business Survey

The aggregate results from the Farm Business Survey are being compared with results from other sources (e.g. the Agricultural Census). This is contributing to the National Statistics review, where there is now a requirement to examine each major area of statistics at least once every five years. It is particularly important to do this in respect of the Farm Business Survey to ensure that there is no bias due to the difficulty of recruiting new participants. A degree of difference is expected, reflecting the fact that the Farm Business Survey is only designed to produce a micro level dataset with key financial totals.

While it is too early to judge the results, there are some encouraging signs. For example, there is a very close match between the full time labour estimated by the FBS and the census. As expected, the FBS gives a higher number of part-time workers, reflecting the yearly pattern, rather than the June census spot date. The cereals area from the FBS is about 12% higher. While the slaughtered sheep and pigs are slightly higher than the official slaughtering figures, the FBS cattle slaughtering figures are lower. More work is needed to look into these differences.

4.3 Example 3: Labour Figures

Comparisons have been made between the employment figures collected in the agricultural census and the estimates from the UK (household) Labour Force Survey conducted by the ONS (Office for National Statistics). If a close match could be obtained with the LFS source, this might help to reduce the labour data collected in the agricultural surveys. While the LFS covers workers in hunting, forestry and fishing, this should largely reflect the trends in agriculture.

Unfortunately the datasets have not proved comparable, particularly over later years. For example, both sources showed that the number of agricultural employees were on a consistent downward trend between 1985 and 1991. Since then the series have diverged. Between 1999 and 2000, for example, the census series showed an 11% decrease (consistent with the fall in production and income) whereas the LFS series showed a rise of 6%. The differences are being pursued with the ONS. The self-employed figures are not comparable.

5. Conclusions

There is scope to use administrative and other statistical sources to improve the quality and value of statistics. This is not easy, costless or straightforward. It can raise more questions than provide solutions. It is preferable for checks to be designed and built into systems and planned in advance. We are working on our vision to make maximum use of the data that is collected, avoiding as much duplication as possible. From the statistical perspective, we are seeking to maximise use of both statistical and administrative datasets.

References

MAFF’s Role in E-Business (September 2000).
http://www.maff.gov.uk/ebus/ebus.htm

Applying Eurostat's Quality Approach to Agricultural Statistics

Werner Grünewald, Rainer Muthmann, Robert Poschacher
Statistical Office of the European Communities (Eurostat)
Jean Monnet Building, L-2920 Luxembourg
Tel.: +352-4301, extension: 33280-W. Grünewald,
37260-R. Muthmann, 35042-R. Poschacher
e-mail: werner.grunewald@cec.eu.int; rainer.muthmann@cec.eu.int;
robert.poschacher@cec.eu.int

1. Introduction and goals

Quality has become a buzzword nowadays. Everybody speaks about quality, nobody is against quality - and some people cannot stand the term any longer. One reason why people become critical about “quality” is the fact that the term is fairly vague, not tangible and perhaps even over-used.
Nevertheless “quality” is a very important issue in the private and public sector. Therefore great attention is paid to the “quality” aspect by the European Commission. The term “quality” is comprehensive and “multifunctional”. In this paper we want to describe how Eurostat applies a quality approach in agricultural statistics. We begin our exposition with the definition of “quality” to create the framework for our further approach. We then describe the general quality approach of Eurostat, which is the basis for the measurement of quality in European statistics in general and also in agricultural statistics (quality reports). Then we describe the implementation of the quality approach in agricultural statistics by giving some concrete examples. Finally we make a summary and give an outlook on quality aspects in agricultural statistics.
So, the issues, which will be dealt with in this paper, are the following:

- What is quality?
- The quality approach of Eurostat
- Measurement of quality in European agricultural statistics
- Implementation of quality in agricultural statistics
- Conclusions and future developmentsDefinition of quality

2. Definition of quality

Official statistics has been increasingly subject in recent years to pressure from “stakeholders” to follow successful paths from the private sector. Official statistics is also expected to produce high-quality products and services.
But when one looks back in history, quality considerations have always been of concern for statistics in general and for statistical services in particular. What is new is, irrespective of the concrete definition of quality applied, the widening of the concept “quality”. In the past quality tended to be a synonym for what might be called

"accuracy" or "reliability". The term "quality" today is broader. Statistical services like Statistics Canada, the CBS in the Netherlands, Statistics Sweden, Statistics New Zealand or the ABS from Australia, just to mention some prominent examples, have elaborated the contents of "statistical quality" in the last decade. They all have influenced Eurostat's definition of quality that forms the basis of its work nowadays.

The starting point of Eurostat's definition of quality of statistics is the ISO norm 8402 that defines quality as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs". This very general definition puts the user and his needs in the centre of all considerations. Looking at the term "quality" from the point of view of a user of international statistics has led Eurostat to a definition of statistical quality composed of seven dimensions:

Relevance

Statistics are relevant when they meet users' needs. Relevance requires the identification of users and their expectations.

Accuracy

Accuracy is defined as the closeness between the estimated value and the (unknown) true value. Assessing the accuracy involves the analysis of the total error associated with the available estimate.

Timeliness and punctuality

Statistics are only useful when the figures are up-to-date and published on time at pre-established dates.

Accessibility and clarity

Data have most value when they are easily accessible by users, available in the form users desire, and are adequately documented ("metadata" according to the type of user). Assistance in using and interpreting figures should be part of the providers' tasks.

Comparability

Data are most useful when they enable reliable comparisons across space like countries or regions and over time.

Coherence

When originating from a single source, statistics are coherent in that elementary concepts can be combined reliably in more complex ways. When originating from diverse sources such as different surveys with differing frequencies, statistics are coherent insofar as they are based on common definitions, classifications and methodological standards.

Completeness

Areas for which statistics are available should reflect the needs and priorities expressed by the users. Completeness is an extension to relevance as completeness does not only mean that statistics should serve user needs but they should serve them as completely as possible, taking restricted resources into account.

It is worth noting that trade-offs exist between the different dimensions. The most obvious example is the trade-off between timeliness and accuracy. Accuracy might go down when timeliness improves. Such contradictions cannot be solved in general terms. Priorities have to be fixed individually depending on the situation. Timeliness might be a priority in some areas like short term business statistics, accuracy in others.

Another remark concerns costs. Eurostat has decided not to include costs as a separate dimension. Though costs certainly influence any of the seven dimensions presented above, they can be more seen as a constraint than as a dimension of quality itself. Quality of data is not higher just because their production costs go down. The reduction

of costs inevitably influences at least one of the seven dimensions of quality in a negative way.
An important prerequisite for good agriculture and environment policies is high quality statistical data in these domains. And significant quality improvements in statistical data are cheap compared to the amount that can be misspent where poor political decisions are taken based on low quality data.

3. The quality approach of Eurostat

The definition of quality of statistical products is just one out of several important elements of Eurostat's general quality approach. The basis of this approach is Eurostat's Corporate Plan stating mission, aims, roles, values, objectives, strengths and weaknesses as well as threats and opportunities of the office. The current version covers the period until the year 2005. Such plans exist at the office level as well as at individual unit level. The user is in the centre of this approach, as can be seen for example from the fact that the first two out of six objectives of the office relate to (the satisfaction of the) users inside and outside the European Commission. Another key issue is the satisfaction and motivation of Eurostat's staff.
Though the users (and the staff) are in the centre of Eurostat's interest, the office must also take into account the view of the producers. Eurostat sees itself as member of a system of statistical authorities of the Member States and the European Union, the European Statistical System (ESS), covering all institutions of the European Union producing official statistics. The ESS therefore includes the National Statistical Institutes but also ministries and other organisations involved in the production of official statistics. This system is governed by the principle of subsidiarity.
Taking all these points into account means for Eurostat in the field of agricultural statistics, as in all other areas, satisfying its (main) clients with high quality products and services while respecting the principle of subsidiarity. Key elements are a satisfied and motivated staff and organisational efficiency particularly in these times of scarce resources. The achievement of these objectives requires appropriate tools. A key aspect are indicators measuring the overall performance of the office with respect to its objectives and the individual performance of single products like data on specific agricultural areas. The development of such indicators is far from being trivial, and their implementation often requires additional collection efforts. They are however a prerequisite to a long term quality assurance of Eurostat's products and services.

4. Measurement of quality in European agricultural statistics (EAS)

4.1 The model quality report

It becomes obvious from the considerations on how to define quality of statistics that quality cannot be measured by just one indicator. The definitions presented do allow one overall indicator to be developed. However, users as well as producers have a strong interest in getting objective and comprehensive information about the different dimensions of quality of individual statistics. In order to collect the necessary information, Eurostat has developed an appropriate tool, the quality report. This report tries to summarise all the relevant characteristics that statistics may have, based on the definition presented above.

Statistics are not homogeneous with respect to Eurostat's seven dimensions of quality. Some statistics are based on censuses, others on sample surveys. Some statistics are based on definitions and concepts harmonised at world or at least European level, other statistics use nationally different concepts and definitions first and are then harmonised at EU-level. Whatever the situation of the individual statistics is, the quality report tool should be able to take all different possibilities into account. Eurostat has therefore developed a model quality report that basically includes all possibilities. This model quality report has to be adapted in all practical applications to the features of the statistics under consideration.

The model quality report identifies different characteristics for each of the seven dimensions. The information is put together in the report and can be used for further analysis such as the calculation of quantitative or qualitative indicators. In order to give an overview on the contents of such a report, the table 1 summarises the information that might be collected in a quality report.

Table 1: *Characteristics of the model quality report*

Dimension of quality	Characteristics of statistics
Relevance	Description of users, user needs, indication of the relevance for users
Accuracy	Sampling error (random vs. non-random survey, variance, coefficient of variation, reliability indicators for non-random surveys), non-sampling error (frame errors, measurement errors, processing errors, non-response errors, modelling errors), qualitative judgement of the overall accuracy
Timeliness and punctuality	Reference period, transmission dates, coverage, dissemination tools, punctuality of the production
Accessibility and clarity	Forms of dissemination, additional documentation, clarity of publications, related information services offered
Comparability	Comparability over time, across space (“geographical comparability”) and between statistical domains with respect to concepts or measurement and their effects
Coherence	Description of the relevant aspects of coherence and their effects
Completeness	Estimation of the total number of statistics requested by the users

4.2 An application of the model quality report in agricultural and fisheries statistics: the quality report on “Fish catch statistics for the Northeast Atlantic”

Any practical application of the model quality report first requires a deep understanding of the product under consideration. It would be here too far reaching to go into the details of the example selected (from fisheries statistics). Basic features of the “Catch statistics for the Northeast Atlantic” however are that

- Eurostat co-operates in this area with ICES and FAO;
- concepts and definitions are harmonised and widely applied by national authorities;

- information on catches of larger vessels (larger than 10 m in length) is collected through administrative procedures;
- information on catches of all other vessels is collected through non-random tools on a sporadic basis;
- available information is annual and can be broken down by species, fishing region and country (flag of vessel).

Applying the model quality report to the area of catch statistics for the Northeast Atlantic leads to a request of information as described in table 2. It shows adaptations to the general framework as well as examples of concrete indicators at the same time.
Table 2: Characteristics and example indicators of the quality report on "Catch statistics for the Northeast Atlantic"

Dimension of quality	**Characteristics of statistics** ***Example indicators***
Relevance	Description of users, user needs, indication of the relevance for users
Accuracy	*Qualitative indicators on core aspects of reliability for non-random surveys with a specific reference to non-sampling errors such as level of coverage or overall quality, or existence of methodological reports*
Timeliness and punctuality	Reference period/ transmission dates/ coverage/ dissemination tools/ *punctuality of the production with qualitative or quantitative indicators for example on the timeliness of transmission, average processing time in Eurostat or on the conformity of transmission to pre-established formats*
Accessibility and clarity	Forms of dissemination, additional documentation, clarity of publications, related information services offered
Comparability	Comparability over time, across space ("geographical comparability") and between statistical domains with respect to concepts or measurement and their effects *with qualitative indicators on the degree of comparability (scale 1 – 5)*
Coherence	Description of the relevant aspects of coherence and their effects
Completeness	Estimation of the total number of statistics requested by the users

4.3 First experience and further use of quality reports: the case of the quality report on "Fish catch statistics for the Northeast Atlantic"

The first version of the quality report on "Catch statistics for the Northeast Atlantic" has been available since November 2000. Though the characteristics requested have been adapted to the needs of catch statistics the first version still contains a fair number of holes. Examples concern indications on the relevance for users, indications on non-sampling errors such as measurement errors and information on the completeness dimension. This situation is typical for all quality reports currently available in Eurostat. It shows that Eurostat's and/or the Member States' knowledge about the different aspects

of quality of statistical data is incomplete. *It is Eurostat's intention to close identified information gaps as far as this is possible under the existing circumstances, including the political necessity*. Each unit in Eurostat concerned has been invited to update their quality reports at least once a year.
Even incomplete quality reports can already be used for concrete improvement actions. The underlying idea is that the quality reports serve as inventories helping all parties involved, to identify those areas where improvement actions seem to be most urgent and/or feasible. The follow-up of the improvement actions is in the hands of the unit concerned, and the outcomes of the actions will be reflected in later versions of the quality reports. Examples of concrete improvement actions that might be based on quality reports like the one on catch statistics for the Northeast Atlantic are presented in the following part of the paper.

5. Implementation of quality

5.1 The FADO-framework

To implement "quality" in the field of agricultural statistics involves Eurostat in a wide range of activities. The overall concept is called "FADO" (Future Agricultural Data Outline). FADO can be seen as the framework, which helps us to face the new challenges and to look at future data needs in the field of agriculture including its interaction with the environment and rural development.
The FADO steering document which is discussed regularly in the different working groups on agriculural statistics comprises two annexes:

- The first annex covers FADO relevant activities within the regular Work Program on Agriculural Statistics, whereas
- The second annex covers activities going beyond the regular Work Program.

In the FADO steering document the work on different issues has either the status of ongoing work, work foreseen (includes in current work plans) or work proposed (under consideration/awaiting resources).
Within the overall framework for our work that FADO gives us, there are various concrete measures by which "quality" is implemented in the field of agricultural statistics.

5.2 TAPAS (Technical Action Plan for Agricultural Statistics)

5.2.1 Introduction and general remarks

On 25th June 1996 the Council adopted Decision 96/411/EC on improving Community agricultural statistics, better known under the name of TAPAS. The aim of this decision is to ensure that Community agricultural statistics meet the needs for information resulting from the reform of the common agricultural policy.
In implementing this Council Decision, the Commission approves annual technical action plans. The basic framework for these technical action plans is the following: DG Agriculture provides financial resources, countries propose actions on predefined fields of agricultural statistics of particular current community interest; Member States cooperate with Eurostat and DG Agriculture on a voluntary basis; work is coordinated

by the Agricultural Statistics Committee and technical and methodical details of actions in the various fields covered are discussed in the corresponding working groups.

5.2.2 Fields of work

Several fields of agricultural statistics have already been improved with the help of TAPAS and some new fields are being started.
The main areas which have been tackled are:

- Early estimates of sowing
- Fruit and vegetable production statistics
- Supply balance sheets (meat balances, crop balances, fodder balances, other balances)
- Use of pesticides
- Animal production forecasts

New fields are:

- Agri-environmental indicators
- Environmental aspects of agricultural accounting

5.2.3 A concrete example: the work on fodder balances

The work on fodder balances gives a good insight into the achievements of the TAPAS work. “Fodder” balance sheets, which should rather be called “animal feed” balance sheets, form part of the statistical programme and the results contained in them are highly important for the Common Agricultural Policy (CAP). The European Commission needs accurate, detailed and methodically well founded feed balances, to be able to take appropriate measures in this field.
The Member States have stepped up their efforts under the TAPAS action plans to improve these statistics. The investigation of the supply and production of animal feed stuffs was as comprehensive as possible, taking into consideration, in collaboration with experts, rather varied sources. Thus, the problems of the complete identification of fodder could be given concrete expression. Very exact descriptions for basic data used were made. With regard to demand, detailed calculation plans were drawn up, taking into account each countries national situation. In addition, an allocation of fodder by categories of animals was made on the basis of the composition of fodder, on the basis of the official definitions of fodder and on the basis of the specialised literature on animal feed.
The principal points which have been worked on concerning the fodder balances are the following:

- Better identification of the supply of fodder (especially of green fodder),
- Data retrieval with regard to the output and the production of meadows,
- Identification of mixed fodder,
- Broad calculations and estimates of demand,
- Allocation of resources by categories of animals,
- Summary and comparison of the various sources of data for drawing-up the fodder balance sheets,
- Use of conversion coefficients which have to be adapted periodically.

5.2.4 Evaluation of TAPAS

A first evaluation of the TAPAS action plans leads to the following conclusions:

- The presence of the TAPAS tool has enabled the Member States to undertake a number of adaptations of their agricultural statistics system, that would not have been possible without a financial contribution, even if only partial, on the part of the Community. The returns from the amounts invested in TAPAS are considerable.
- The approach which has been followed up to now in establishing the technical action plans, to give priority to the Community needs in the definition of the priority fields, has avoided the dispersal of the financial resources on a too high number of actions, which would have had a lower added value on the Community level.
- The implementation of TAPAS has contributed, and should continue to contribute, to the reorientation of the system of agricultural statistics towards the practical needs of the users at the Community and national level.

5.3 Introduction of estimates and forecasts

Another aspect of "quality in agricultural statistics" is the ability to offer the users complete time series without missing data. Often in agricultural statistics this is a serious problem. If one takes for example a look at the pig surveys, one sees the problem: Not all member countries deliver the data at the same time, since the legislation of the EU takes special interests of the Member States into account. In addition to that, Eurostat makes use of the "Agriflex principle" which means that not all Member States must follow the given requirements in the same way. Example: If a certain sector of agricultural production in a country is not very important, the amount of standard data demanded can be reduced and/or eliminated.

Therefore we have on the one hand the situation that - because of different timings of pig surveys due the legislation of the EU and to the Agriflex principle - in nine out of twelve months somewhere in the EU pig surveys are carried out. On the other hand Eurostat has to offer its users at least three times per year an EU-15 aggregate of the pig livestock which can only be reliably calculated if one has at least 85 % of the population of the previous period. To solve this problem, Eurostat applies „time series modelling". With this method we calculate the missing values and we use it also to make short-term forecasts. We chose this method, because our experience showed that the calculated values are of high and considerable quality. This approach was also discussed with EU Member States in our working group "Animal Products Statistics". There it was agreed that this method is appropriate to solve the problem of the missing values, but some Member States do not want the figures for their country to be published. As a result of this discussion, Eurostat calculates the missing values, but only publishes figures for the EU-15-aggregate and not on the level of the Member States.

5.4 The use of administrative sources in agricultural statistics

Where data are available from other sources than statistical surveys it is a natural reaction to try to use them and to avoid apparently double work for respondents and for administrations. This is not without problems however, especially in a context of international comparisons.

While statistical surveys have been developed to provide a certain level of precision and comparability, the use of data from other sources as a substitute for statistical data has three main disadvantages. The first is that the data are not collected for the information

purposes of the statistical survey. The observation units, total statistical population, concepts and frequency are not chosen having this in mind. Secondly the respondents usually have an interest in colouring their answers in relation to the context in which they are given. Thirdly comparability over time and from country to country can be damaged by changes and differences in the operation of the data collection.
To make sure that the use of administrative data does not lead to a decline in quality, Eurostat applies a range of measures to ensure the quality of agricultural statistics when using administrative sources. These basic criteria, to be taken into account when considering the use of administrative data for statistical purposes, are:

- The statistical use of an administrative data bank (ADB) should preferably be planned ex ante in close cooperation between statisticians and the main user of the ADB.
- It should be made clear which elements of the statistical population could be taken from the ADB. It should be proved that the ADB is comprehensive and representative as to these elements.
- Moreover - and that is of the utmost importance - the updating procedure of the ADB must assure that the ADB remains comprehensive and representative for the elements which it contains also in the future.
- The remaining elements of the statistical population which cannot be provided via the ADB have to be surveyed in an appropriate way, for example by an adequate sample survey.
- The use of an ADB for statistical purposes should be limited over time and reviewed at regular intervals (every second to third year); a review is always necessary when the underlying rules for building up the ADB are thoroughly revised.
- A regular report to the Agricultural Statistics Committee on the use of ADB for statistical purposes in a specific sector in a Member State increases transparency. In years when a structure survey of agricultural holdings is taking place, a comparison with its results could be of special importance.

To give a concrete example how these basic criteria are taken into account, we use the register of bovine animals in Austria. Austria was authorised to make use of the register of bovine animals to partially replace surveys of bovine livestock (see Commission Decision No 2000/554/EC of 6 September 2000). The basic criteria are concretely taken into account as follows:

- In view of the current level of coverage and representativeness of the operational data base for bovine animals in Austria, it may be used to obtain statistical data on some elements.
- Since not all elements cannot be obtained from the bovine database, they are the subject of supplementary statistical surveys.
- For comparison reasons all the elements have also to be the subject of this supplementary statistical survey. Therefore one can compare the results of the statistical survey with the results of the register which gives some good hints of the quality of the bovine database.
- A cooperation group made up of the responsible officials of “Statistik Austria”, the Federal Ministry for Agriculture (the chief administrator of the bovine database), external representatives of the chambers for agriculture, experts of the university of

agricultural sciences (statistics, cattle breeding) and a representative of the European Commission has been set up to supervise the statistical use of the administrative data base. In particular, this group ensures that the updating procedure of the bovine database is monitored in regular intervals, thus guaranteeing coverage and representativeness of the bovine database. This group will undertake a detailed examination each time the way the bovine database is operating, changes significantly.

- An annual report is to be compiled by Austria on the use of the bovine database for statistical purposes. After the authorisation for the use of the bovine database has expired, Austria has to forward a report to the European Commission, which enables a comparison of the bovine database results with the results of a large sample survey on cattle.

With these measures it can be ensured that the statistical information which is taken from the bovine database in Austria, is of high quality and helps the Commission to implement the CAP. In addition, it is assured that the response burden of farmers is minimized as far as data for administrative and statistical purposes are concerned. It can be foreseen that other countries will also use administrative sources in the future to replace statistical surveys and therefore this approach can also be a model for other countries and also for other statistical areas.

5.5 Statistics in Focus

To provide the decision makers or media with information of a high quality in an easy way and which helps them in their daily work, Eurostat produces the "Statistics in Focus" (SIF). The aim is to provide the users with a user-friendly, harmonised, stable, documented, up to date, cross domain off line and on line application. The SIF's are grouped into a number of broad themes (theme 5 is agriculture and fisheries) and are of four or eight pages in size. They are available in German, English and French. SIF became one of the most successful Eurostat products with 7.400 privileged subscribers and 9.130 paying subscribers. SIF is the most consulted collection on the Eurostat web site (http://europa.eu.int/eurostat.html) with 50% more consultations than news releases (which are free).

6. Outlook

Various aspects of quality become more and more important in agricultural statistics. Eurostat tries to fulfil the users requirements with many different measures. But implementing these measures and ensuring their application also requires financial resources. Therefore a fundamental question is: are the users willing to pay more for statistics which is of a very high quality and if so, how much more? This is not only a basic, but also a very difficult question to answer.
Agricultural statistics have shared the difficulties afflicting statistics generally. Statistics are frequently seen as a burden on respondents, as a brake on the efficiency of the units or as an unwarranted intrusion into private life. Administrative systems have been set up with information aspects which overlap traditional statistics. Some sources such as customs data have been at least partially removed. Statistical administrations may see the way forward in generally reducing the scope of official statistics. Such decisions,

however, neglect the opportunity costs of statistics, i.e. the costs of bad decisions taken in the absence of data of sufficient quality. Some countries have experiences along this line and, having found they had reduced their national statistics too much, they had to build them up again. In a world which is increasingly flooded with information much of it of unknown quality, there is a clear need for the neutral, authoritative reference that official statistics provides and which is available and easily accessible for the public.

It is certainly true that official statistics should not try to meet all needs. One can admit that official statistical services are not usually very strong at providing daily or ad hoc data for short-term or real time reactions. There should be a place for studies, ad hoc surveys and panels, from private as well as public sources, to meet special demands. However, these ad-hoc needs need to be embedded into a framework of regular, official key figures. In this way, overall adequate information is made available, not only for central policy fields but for informed debate and efficient economic reactions. For cost efficiency considerations official statistics often aim at information systems meeting multiple needs and providing objective reference results generally in the form of long series. These systems depend on individual statistical actions being embedded in a framework of common standards, methodologies, classifications and accounting structures, ensuring comparability between data from different actions, reuse of data collected in one context for other purposes and inter country comparability. Eurostat and the national statistical services of the Union Member States also cooperate with international organisations like the UN (ECE), FAO or OECD to ensure maximum comparability also outside the EU.

To develop and establish a statistical system one needs highly qualified experts, a motivated staff, but also the necessary time and experience. The European Statistical System (ESS) is a fairly developed and very complex example of such a system. However, despite all efforts in the past, there is still enough room for improvement. The ESS recently analysed the current strengths and weaknesses of the system the results of which were presented at the international conference on quality in official statistics in Stockholm in May. It would be too far reaching to summarise all findings here. One conclusion was that not all strengths and weaknesses identified have the same importance for all statistical areas. Looking at quality in the domain of agricultural and environmental statistics, the following aspects seem particularly relevant:

- Quality thinking must be intensified in the future;
- Quality requirements are not sufficiently operationalised and quality measurements have still to be improved considerably;
- Quality of data at European level depends on the quality of the data delivered by the producers in the Member States;
- Availability of harmonised meta data is still the exception and not the rule.

The ESS is a dynamic and flexible system which can be adapted and modified if one faces new conditions such as the enlargement of the EU. This system can perhaps serve others as an example of “good practice”, but the ESS can, of course, also learn from statistical work in other international associations.

Standards for Official Statistics

Robin Slater
Australian Bureau of Statistics
PO Box 10, BELCONNEN ACT 2616, AUSTRALIA
Tel.: +61 2 62525166, Fax: +61 2 62531095, email: robin.slater@abs.gov.au

Abstract: Official agriculture statistics require a range of standards in order to provide maximum benefit to users. This paper describes key standards applying to agriculture statistics produced by the Australian Bureau of Statistics. The focus here is on conceptual, process related and dissemination standards.

Keywords: agriculture, statistics, standards, Australia

1. Introduction

This paper presents an overview of the system of standards and related processes that apply to the production of official agricultural statistics in Australia.
It is not intended as a normative statement of standards that should apply in other countries - FAO and Eurostat are in a better position to look at agriculture standards from an international perspective, not to mention the need for individual agencies to look first and foremost to the needs of their own citizenry. Nor is it necessarily "best practice" in an international sense, as the set of standards used in the Australian Bureau of Statistics (ABS) reflect in part our own history and circumstances. As a consequence the context in which agricultural statistics in Australia are produced, and some explanation of what we mean by "official statistics" is appropriate.

1.1 Why Standards?

The Macquarie Dictionary gives several definitions of the word "standard", the most relevant to our context being:

- anything taken by general consent as a basis of comparison; an approved model
- a level of quality which is regarded as normal, adequate, or acceptable.

A Mac Donalds hamburger, the Microsoft Windows operating system and accounting standards for business reporting are but a few illustrations of how standards add value to goods and services. What do we gain from having standards for official statistics?
Put simply, standards assist in maximising the <u>effectiveness</u> of statistical outputs (especially in terms of utility to the user) and the <u>efficiency</u> of the production process.

For the purposes of this paper, I define effectiveness in terms of support for comparability (over time, space, industry, etc) and coherence (i.e., the capacity for integration) of the official statistics. While comparability and coherence are important for any dataset, they are particularly important where data are obtained from multiple sources and have to be combined in some way, or where the outputs are used in a wide variety of contexts. This is a common situation confronting official statisticians, such as when compiling national accounts. Standards for these aspects of effectiveness in turn promote transparency and hence integrity. Arguably, integrity is the most important quality attribute that a national statistical office can have.

While the quality dimension of official statistics is most directly related to effectiveness, the ultimate degree of quality that can be attained, given finite resources, is indirectly dependent on the efficiency of the process of producing these statistics. Thus from the perspective of quality, standards relating to the process can be as important as those related to the product. This paper aims to address both.

Paradoxically, standards can also enhance flexibility, both operationally and in terms of the variety of uses to which statistics can be applied. Even so, standards for official statistics, particularly those focussing on the dimension of quality, cannot be viewed in isolation from intended purpose. Another factor to bear in mind is that standards can have both positive and negative synergistic effects - eg, accounting and software standards tend to complement each other, while accountability and secrecy standards are often in conflict.

There are many categories of standards that should, and do, apply to official statistics. These include those for accountability (to government, users and providers), concepts, presentation of statistical output, financial management, people management, data transmission, and so on. Not all of these are obviously related to the quality of the statistics, but most do have some ultimate bearing on quality. This paper will address a core subset of these standards, **focussing on standards related to concepts, process control and dissemination.**

Finally, the boundary between standards, guidelines and good practice are often blurred. The material described here touches on all three, but the focus is on standards.

1.2 What Are "Official Statistics"?

The UN Statistical Commission adopted a number a fundamental principles of official statistics at its meeting in April 1994. The resolution begins with the statement:

> "Official statistics provide an indispensable element in the information system of a democratic society, servicing the government, the economy, and the public with data about the economic, demographic, social, and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information."

Further useful statements on the nature of official statistics come from the 1993 United Kingdom White Paper on Open Government; key statements include:

> "Official statistics are collected by government to inform debate, decision making and research both within government and by the wider community."
>
> "Reliable social and economic statistics are fundamental to ... open government (and) it is the responsibility of government to provide them and to maintain public confidence in them."

These attributes of official statistics are embodied in the mission statement of the Australian Bureau of Statistics, Australia's national statistical office:

> "We assist and encourage informed decision-making, research and discussion within governments and the community, by providing a high quality, objective and responsive national statistical service."

1.3 The ABS Context

These statements define official statistics well in the broad sense, as regards who produces them, and how important it is to achieve these aims.

Australia has a centralised national statistical service - i.e., we have a single agency, the ABS, which provides the majority of official statistics for all levels of government (federal, state/territory and local) and supports many portfolio agencies. The mix, quantity and quality of the agricultural statistics we produce at ABS need to be seen in this context. For example, our agriculture statistics program must compete for ABS resources against other statistical programs in the ABS, such as manufacturing, national accounts, labour statistics, and so on. Similarly, the agriculture program has adopted many standards that apply to ABS collections in general, not just to the agriculture program.

Increased standardisation of concepts and operations has been a long standing theme in ABS economic statistics. In the late 1960s and 70s, integration of business surveys based on a shared business register, and adoption of the UN System of National Accounts (SNA) concepts, characterised the experience of a generation of ABS managers. In the 1980s, the introduction of form (questionnaire) design standards and the move to centralise collection activity for individual collections marked another major shift to standardising operations within collections. In the 1990s, standardised practices for sample and frame maintenance procedures (SFMP), generalised software utilities for statistical processing, a corporate output data warehouse and upgraded publication standards continued the trend of increasing standardisation.

The next decade is shaping up as just as challenging, with a strong emphasis on further standardisation. One particular challenge is the increasing need to apply ABS standards to data sources it cannot directly control. Despite having a centralised national statistical office, over the years the concept of "official statistics" in Australia has tended to broaden beyond the limit of data collected directly by the ABS. Increasingly, administrative datasets are seen as essential elements of our national statistical system, primarily because they can extend the range of statistical data we can provide at low cost. Examples of this in the agriculture field include our increasing use of taxation data for frames, sample design and ultimately data substitution, and use of farm surveys from the Australian Bureau of Agricultural and Resource Economics (ABARE).

1.4. Legislative Framework

In the private sector, legislation typically constrains what organisations can do in the conduct of their business; in the public sector, legislation enables organisations to conduct particular functions, often with quite specific provisions for the powers that an organisation can exercise and how. An explanation of standards related to official statistics must logically begin with the legislative foundation upon which the national statistical system exists.

In Australia, the *Census and Statistics Act 1905* provides the Australian Statistician with the authority to conduct statistical collections, including the Census of Population and Housing and, when necessary, to direct a person to provide statistical information. The Act requires the ABS to publish and disseminate compilations and analyses of statistical information and to maintain the secrecy of information collected under the Act.

This Act, and related legislation, define the function, powers and responsibilities of the ABS. Legislation covers issues such as authority to collect data, the need to table any new collections in the Parliament, authorised officers, confidentiality and secrecy, etc. Other legislation deals with more general obligations on Commonwealth (i.e. federal) agencies, such as freedom of information, archiving, privacy, merit protection, etc. It makes little sense to talk about standards for official statistics, whether related to quality or not, without acknowledging the importance of relevant legislation to the mandate of the statistical agency.

2. Conceptual standards

The ABS has developed a framework for Integrated Economic Statistics under which most of our business surveys are conducted. This framework requires that the statistical structure of each business entity is based on a standard units model, and an industry code is allocated based on predominant activity. A key component of the framework is a centralised business register (Inteframe) which stores this information about each business and which is used to produce population frames for collections of economic statistics, based on the industry code.

A set of standard classifications is used to describe characteristics of businesses, such as their industry, size, type of legal organisation, and the commodities they produce. In addition, standard data item definitions have been developed, and these are accompanied by the use of standard questions in questionnaires.

The ABS has two primary survey vehicles as the basis of the agricultural statistics we produce.
The Agricultural Finance Survey (AFS) is part of a series of annual economic collections, and is subject to all of the standards imposed by the Integrated Economic Statistics Framework. Its framework is limited to those units which are coded to the Agriculture Industry, and its results can be directly compared with those from other industry collections.
The Agriculture Census and its related surveys are outside of this framework, since they are based on "activity" rather than "industry". While Inteframe is still used to generate a population frame, this frame comprises all businesses with agricultural activity. This can include businesses whose main activity is not agriculture, and thus these collections are not restricted to units classified to the agriculture industry.

2.1 Units

The current ABS statistical units model is based on a combination of "real world" units (i.e. legal and physical) and management accounting units. The model is applied to all businesses held on the ABS business register, Inteframe, and is applied to all collections taking their frame from the register (such as the AFS). The current model has five levels:

(i) Enterprise Group (representing the highest level of common ownership)
(ii) Enterprise (legal entity)
(iii) Management Unit (the highest level management accounting unit within an Enterprise Group, able to support 2-digit industry homogeneity)
(iv) Establishment (a lower level accounting unit able to provide detailed revenue, stocks, employment, etc data; The Establishment is bounded by State/Territory and supports 4-digit industry homogeneity)
(v) Location (the physical location(s) relating to the establishment).

For Agriculture statistics, the relevant units are the Management Unit and the Establishment. The Management Unit is targeted by the AFS to collect the detailed farm business performance ("financing") information. Agricultural commodity data (e.g. in the Agriculture Census) is collected from a unit conceptually identified with the Establishment. The "establishment" for agriculture collections is usually based on the farm location or a grouping of farm locations bounded by a specified local geographical area, rather than State/Territory, thus supporting greater geographic detail in the aggregates.
Increasingly, the traditional boundaries of farming activity are becoming blurred by the combination of farmers diversifying into non-farm activities and value adding processes attached to agricultural production. Where significant value added processes are undertaken in conjunction with agricultural activity, such as the farming of chickens and poultry processing, the ABS endeavours to split the Management Unit along industry lines.

Over the next few years, the ABS will be exploiting recent reforms to the Australian taxation system to its advantage. First, the introduction for tax purposes of a central government register of all Australian businesses, including the adoption of a standard unique identifier - the Australian Business Number or ABN - has major implications for official statistics. In particular, having a reliable and universally used identifier for all businesses will improve the quality and efficiency of our frames and sample designs. Over the longer term, we expect to be able to use data collected by the Australian Taxation Office to supplement the data collected in our surveys.

The future for the units model used in agriculture is further complicated by our long term ambition to adopt better spatial units and to link farms with households - e.g., so data for instance, from the agriculture census and the population census can be integrated. Our agriculture units model is ultimately likely to be composed of farm businesses (equivalent to the current agriculture "establishment"), ABN units (legal entities for tax purposes), farm households, and land holdings (i.e., the physical parcels of land defined by cadastral boundaries).

2.2 Classifications

Australia bases its major classifications on international standards such as the International Standard Industrial Classification of All Economic Activities (ISIC) and the Central Product Classification (CPC). ABS economic classifications align as closely as possible to the international standards, and use the same underpinning concepts. The Australian and New Zealand Standard Industrial Classification (ANZSIC) is derived from the ISIC and closely aligns for agriculture, ANZSIC being more detailed to cater for local activities. Whilst the CPC forms the basis of the Australian and New Zealand Standard Commodity Classification (ANZSCC), aligning at the three to five digit level of the CPC, an extension of this classification has been developed for agriculture to incorporate data requirements such as area of holdings by type of crop, areas under cultivation/tree plantings and production forecasting information.

Interest in agricultural statistics is broadening from the traditional farm performance and production data to include information on the farm household, environmental issues and non-agriculture farm activities such as value adding and marketing activities. Where possible, data are collected based on relevant international classifications.

For many users of data the concept of "agriculture" as an industry can no longer be restricted just to farm production. They want to expand the scope of the industry definition to include not only on farm value adding but also the manufacturing and marketing processes related to it as well - the "farm gate to plate" concept. ABS is endeavouring to meet this need by developing a classification based on the standard industry and commodity classifications that includes all stages of production, from planting through to the final packaged product. Currently ABS does not collect data to populate all of this alternate view of the industry; however it is hoped that future ABS data collection and data from non-ABS sources will provide reasonable coverage.

2.3 Data Items

Data Items in ABS economic collections are developed in an integrated manner. Those leading to the derivation of business performance indicators, in particular, are collected in an identical manner in all economic business collections.

Whilst the National Accounts sets the framework for what data *should* be collected from businesses, Australian accounting standards, Corporation Law and taxation reporting requirements form the basis of *how* data is collected from business management accounts. Data item definitions change in line with any changes to these external standards. New data items are tested extensively before being implemented in a collection and their performance measured afterward by post enumeration studies. To ensure integration and comparability of data across business collections, a core set of data items and accompanying collection question wording are developed each year for inclusion in all relevant surveys. This core set of data items and associated question wording must form the basis of any survey content where financial data is to be collected.

The development of other data items and question wording, e.g. commodity or survey specific data items, follows a strict process of development. Question wording is first developed based on the relevant standard classification, then field tested and submitted for approval, prior to implementation. Where the same questions are likely to be included in more than one industry collection, they are tested across industries before inclusion.

3. Process related standards

As the official statistical agency for Australia, the ABS has actively pursued many avenues that improve the quality of its statistics. These quality improvement efforts focus on the process of producing statistics as often as the statistics themselves.

In the mid-1980s, the ABS embarked upon a major upgrade of its questionnaires. This involved developing, in conjunction with the Communications Research Institute of Australia (CRIA), a set of design standards and procedures that have controlled the "look and feel" of ABS forms. Even more importantly, the standards extended to how questionnaires are developed and tested.

The data items that appear on forms as questions undergo an internal approval process to ensure consistent use of terms and definitions across different collections. Standard Sample and Frame Maintenance Procedures likewise provide for the consistent treatment of units within and across collections, with specified procedures for dealing with non-reporting, partially-reporting, new and defunct units.

The timeliness of ABS agriculture statistics releases is determined primarily by the need to synchronise publication with major external events, such as the setting of State budgets and the annual commodity forecasting conference, *Outlook*. The resulting timeliness is generally better than that achieved for other ABS business based collections. The same applies to survey response rates, which in regard to ABS agriculture collections are of the order of 90-98% of live units.

Nearly every dimension of statistical activity, from the transformation of client requirements into collection specifications to the final dissemination of data, is subject to corporately maintained standards to some degree. Typically, these standards are either embedded in generalised software tools or in standardised business processes (forms design, frame maintenance procedures).

3.1 Statistical Methodology

Sample stratification in ABS business based surveys, including agriculture, is typically by industry, size and geography, and standard software facilities are often used.

Statistical methodology during processing is becoming increasingly regulated, mainly through the introduction of generalised software tools such as those for estimation (GenEst, which uses design-based estimation methodology and produces Jack-knife variances and RSEs), imputation (GenImp) and windsorisation (GenWins). These utilities rely on a range of Unit Status Indicators (USIs), standard codes that summarise a range of information about the processing status of each record in the survey data file. These standard processing utilities and the re-engineered business processes that flow from using them help to remove the ad hoc clerical intervention that occurred in the past. This approach is fundamental to the ABS' desire to produce a set of integrated statistics across a broad range of subject areas.

A standard set of statistical quality indicators is applied to ABS collections, addressing quality at both the collection level and at estimate level. These measures are derived from the USI codes. They provide standard measures across collections of key quality attributes. They include measures such as response rates, imputation rates, the extent of outliering, extent and type of frame changes (business takeovers etc), use of proxy records, and so on.

Recognising that methodological standards can change, the ABS is progressively moving towards more metadata controlled systems rather than embedding standards in "hard coded" software. Key processes, such as imputation and estimation, are driven by a set of "rules" that are broadly uniform from collection to collection, but which can change. If a change does occur, the metadata need only be modified once, benefiting all collection areas.

3.2 Provider Management

In recent years, aware of the reporting load its collections place on the business community, the ABS has moved to standardise its approach to measuring and reducing "provider load".

This is effected through three primary mechanisms: the measurement of provider load in all business based collections, the survey vetting processes of the Statistical Clearing House, and the introduction of a Business Surveys Charter.

The Statistical Clearing House (SCH) was set up within ABS on 1 July 1997. Its mandate is to vet all surveys conducted by federal government agencies (including the ABS) involving more than 50 respondents. Its aim is to reduce the burden placed by the Commonwealth Government on businesses, particularly small businesses, by eliminating duplication and ensuring the design and conduct of business surveys follow good practices. Typically, the SCH reviews about 150 surveys a year, about a quarter of which result in some form of intervention.

The ABS has been monitoring the total annual load imposed by its surveys on providers for several years. A standard question asking providers to estimate the time taken to complete the questionnaire is applied to all business surveys. The information is used to monitor provider load and is reviewed annually by an executive level Provider Load Committee. This information is used to compile a de facto budget for provider load across ABS business surveys. This process, backed up by the work of the SCH, has been very effective: over the past 5 years, aggregate provider load has been reduced by 37% for all businesses, and 43% for the small business sector.

The Agriculture program in ABS has made a significant contribution to reducing overall ABS provider load on two fronts:

(a) The frequency of the Agricultural Commodity Census has been reduced from annual to five-yearly, with sample surveys in intervening years; and
(b) rationalisation of effort between ABS and ABARE, with a view to completely eliminating duplication between these agencies and increasing data sharing.

The ABS has also introduced a "Business Surveys Charter" which provides information about the ABS and sets out the rights and obligations of businesses which provide statistical information. The Charter explains how businesses can seek help from the ABS, and provides the opportunity for small businesses (those with less than 20 employees) to request exemption where they have been included in a survey for more than three consecutive years. An exemption can be granted in cases where doing so does not significantly bias the results of the survey.

3.3 Form Design and Testing

Like publications, the questionnaires that a statistical agency uses to collect data have a major influence on the corporate image the agency projects to the society it serves. Moreover, these critical collection tools justify the highest standards in communication and presentation to maximise the prospect of a complete and accurate response.

In the ABS, questionnaires for business collections are subject to a formal approval process after being constructed and tested. Three central areas are involved:

- Forms Consultancy Group (providing expert advice on design and evaluation of data collection instruments, whether paper or electronic)
- Economic Standards Section (who monitor and advise on economic and accounting standards, classifications, concepts and international standards)
- the Statistical Clearing House (eliminating duplication between business surveys and ensuring their design and conduct meet acceptable professional standards).

If the questionnaire is to use optical character recognition (OCR) for data capture - a common situation - then a fourth area, the Forms Handling Unit, is consulted to ensure that presentation and printing standards necessary for efficient mechanical reading are in place.

Most business based questionnaires in the ABS change only marginally between iterations of a collection. Moreover, most of the questions used are based on standard concepts, even if the question itself has changed. As a result, form design tends to change in an evolutionary rather than revolutionary way. However, there are exceptions, such as emerging interest in information technology and environmental management. These new areas of interest place more demand on research and development aimed at defining basic concepts and arriving at collection forms which accurately capture them.

Form Development: Before a questionnaire can even be developed, a high level approval and review process is undertaken for all business based surveys. This process is based on an initial strategy paper for the survey, updated regularly. These strategy papers have to address all the key issues relating to the survey, including frequency, purpose of the collection, timetable, user needs, statistical units, data items, classifications, outputs required, the target population, provider load and so on.

In addition to this high level approval process, each questionnaire must also go through several other stages of approval. Critical steps that must be undertaken include:

(i) consult with central standards areas - Forms Consultancy Group; Economic Standards Section; Statistical Clearing House; Forms Handling Unit
(ii) develop and evaluate of a Schedule of Questions
(iii) develop prototype questionnaire
(iv) evaluate the prototype (iteratively evaluate and modify the prototype form using error analysis, observational studies, interviews, respondent visits)
(v) consult with operations areas (with office staff likely to be involved in the data capture, processing or editing)
(vi) pilot test as needed.

In practice the experience and views of the survey area staff, the size or novelty of any changes and timing or resource constraints result in this process being conducted at varying levels of formality. The constant components are in-office research, development and consultations, and respondent visits and testing (discussed below).

Design: ABS business forms are constructed using Adobe Pagemaker, a desktop publishing application. Staff constructing or modifying forms do so in line with the standards set out in the ABS Form Design and Development Manual which prescribes many details of the instrument. They also use standardised page layout templates, graphic objects and common form elements such as the front page layout with its mandatory wording, as well as the margins, fonts, spacing, size and type of data entry boxes, etc, in the body of forms. Where appropriate, page templates also include mandatory questions and wording for national accounts data items.

Testing: While observational studies of respondent behaviour are the ideal method of testing forms, they are usually impractical due to the elapsed time taken to complete a questionnaire. The nature of questionnaire testing is therefore highly dependent on the aspect of the form being examined; however, it generally involves providing draft forms to respondents for their comments, queries or suggestions, then using personal visits or telephone interviews to examine particular concepts, definitions or questions of interest, as well as identified or potential confusion or ambiguity.
The questionnaires generally emerge through iterative development, testing and discussions. Much testing is survey or subject matter specific, though survey areas coordinate the examination and testing of common national accounts related definitions and the resulting standard question wording across surveys. Recent examples include wording related to wages and salaries and goods and services tax.

Approval: After testing, electronic images of forms are lodged on a database accessible via the corporate intranet (based on Lotus Notes). Here stakeholders - primarily the three central areas mentioned in the introduction - provide final comment on the design. By this stage in the process there should not be many substantive issues left to resolve. The Economic Standards area and senior subject matter managers review the form and its development and, if satisfied, formally sign it off (electronically) before forms are printed.

4. Dissemination standards

Once statistical processing has been completed, macrodata datasets are exported to the ABSDB, the ABS output data warehouse. The ABSDB is an important tool for standardising the output of collection areas, and provides a common platform for the generation of survey outputs, whether paper or electronic.
The warehouse includes generalised software tools to help prevent the release of confidential data, and to perform seasonal adjustment, among others. In addition, publication of outputs occurs via another generalised software tool, the Publication Production Workbench (PPW). PPW incorporates the wide array of ABS publication standards supported by extensive metadata and rules for their application. This removes the need for subject matter statisticians to become experts in ABS publishing standards.
Printed publications are still the main method of disseminating data from ABS collections, though increasingly a variety of other media are also used, particularly those based on the ABS web site. In addition to extensive standards for presentation (not addressed here), the ABS has developed a set of standards for release of data, which comprise the following major elements:

(i) Metadata: Data are not generally released without accompanying metadata or reference to such metadata. These metadata generally include some information about the collection vehicle, the main data item definitions, and comparability of data over time. In addition, information on collection methods, standard classifications and methods and directories are published as appropriate, independently of statistical data.

(ii) Confidentiality: The ABS has standard confidentiality ("disclosure avoidance") rules which it applies to released data, and has standard methods of indicating confidentialised cells within tables. Agriculture is a rare exception to the general rule however, with a Ministerial Determination that allows the release of certain agricultural information (primarily commodity data) even when an individual farm may be identified, unless the farm operator requests the data to be suppressed.

(iii) Quality Measures: A number of standard quality measures are used, both at collection level and at estimate level. At collection level, response rates are used to give a general idea of quality. At estimate level, imputation rates and relative standard errors (RSEs) are used to indicate quality. Generally, RSEs are not provided for each estimate, but a table is provided as part of the metadata. Non-sampling error is also acknowledged where relevant, but it is rarely quantified.

The ABS has strict standards regarding the timing of release of its statistics. Data are typically made available, in both paper and electronic form, at 11:30 am on the day of release. This way, no one (whether in government or business) can derive an unfair advantage form early and exclusive access to official statistics. Furthermore, release dates are advertised well in advance - typically several months before the release.

5. Conclusion

The ABS has long held the view that standards are essential to the production of good official statistics, including agriculture statistics. To be cost-effective, standards need to be applied to the process of producing these statistics, not just to the statistics themselves. The experience at ABS is that the degree of standardisation - both across ABS collections and internationally - will continue to increase over time as we get better at mastering complexity and strive for more efficient operations.

References

FAO (1996), *Conducting agricultural censuses and surveys*, No. 6, FAO Statistical Development Series.

UN Statistical Commission (1994), "Fundamental Principles of Official Statistics", *Official Records of the Economic and Social Council*, 1994, Supplement No. 9

"United Kingdom White Paper on Open Government", *Open Government* (July 1993), pp 12-13

Eurostat/IMF/OECD/UN (1993), *System of National Accounts 1993*

Australian Bureau of Statistics (1993), *Australian and New Zealand Standard Industrial Classification*, 1993 Edition

Australian Bureau of Statistics (1996), *Australian and New Zealand Standard Commodity Classification*, Version 1.0

The Importance of Agriculture to a National Economy

- A Methodological and Empirical Analysis Based on German Data -

Ursula Huber, Volker Appel
Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft
Postfach 14 02 70, D-53107 Bonn, Germany
e-mail: volker.appel@bmvel.bund.de

Keywords: agri-business, agriculture, national accounts, value added

1. Introduction

National accounts provide us with information about the structure and development of a country's economy. Moreover, they explain to us how the national income has been generated and used. For this purpose, statistical units are defined and broken down in sectors and sub-sectors. Within these sectors, industries engaged in the same or similar activities are differentiated from each other. In the EU the respective economic units on the lowest structural level are based on a (four-digit) category of the ***N**omenclature générale des **a**ctivités économiques dans les **C**ommunautés **e**uropéennes* (NACE Rev. 1)[1]. The German *Systematik der Wirtschaftszweige, 1993 edition* (WZ 93) is compatible with this.
The classification of the economic sectors subdivides agriculture in agriculture and forestry as well as in fisheries and fish farming. Thus, within the framework of national accounts, data relating to agriculture only identify output in crop farming, livestock farming, commercial hunting, forestry, fisheries, pond farming and fish farming as well as services rendered at the levels of agricultural and forestry production. However, this only reflects the primary production of agricultural holdings, presenting it as an economic sector in its own right. For instance, this economic sector does not include the (commercial) production of foodstuffs, fertilisers and agricultural machinery, nor the processing of products coming from agricultural enterprises and derived by slaughtering, milling, and preserving. According to the systematics of national accounts, this type of activities is included in other economic sectors. However, the importance of agriculture outside the primary production sector is not shown by official national economy indicators.
This limitation does not take into account the actual importance of agriculture in a national economy, particularly since agriculture is closely linked with the upstream and downstream economic sectors. Moreover, the development of upstream and downstream economic sectors has been based on agricultural activities, and their structural development is closely connected to the development of agriculture.
Therefore we will describe a methodological approach which, deviating from the procedures of national accounts, will illustrate not only the importance of agriculture but also the importance of the whole so-called agri-business, that is agriculture

[1] NACE Rev. 1: "Statistical classification of economic activities in the European Community" pursuant to Council Regulation (EEC) No 3037/90 of 9 October 1990.

including its direct upstream and downstream economic sectors. In order to quantify the importance of agri-business for a national economy, the existing concept of national accounts forms the basis for the calculation of the indicators relevant to agri-business. In doing so, existing official statistics will be used. If they are not directly applicable we will show ways of quantifying and estimating data.

2. Definition of "agri-business"

"Agri-business" is a term which is frequently used, but often it has various meanings. On the one hand, the term can mean only profit-oriented sectors upstream of farming. On the other hand, the term can mean economic sectors upstream and downstream of agriculture. In an even broader sense of the word the term "agri-business" can mean agriculture including its upstream and downstream sectors. In this context, in some cases the term "agriculture" means "agriculture and forestry including fisheries", in other cases the term does not have this broad meaning and excludes forestry and fisheries.

Furthermore, it is necessary to be more precise, as e.g. even the group of economic sectors upstream of agriculture either only reflects the sectors producing their products specifically for agriculture or whose products are almost exclusively used in the agricultural sector, or – in a much broader sense – those sectors producing products which are used in the agricultural sector but which can be used in other sectors, too. Examples of the first case, recording only agriculture-related products, are tractor, fertiliser or pesticide industries. In the second case, i.e. products used in agriculture without having been exclusively produced for the agricultural sector, are, for instance, energy and water or tools and equipment (computers) for general use. Agriculture might be an important buyer of these products, but usually it is not the main buyer or the sole target group.

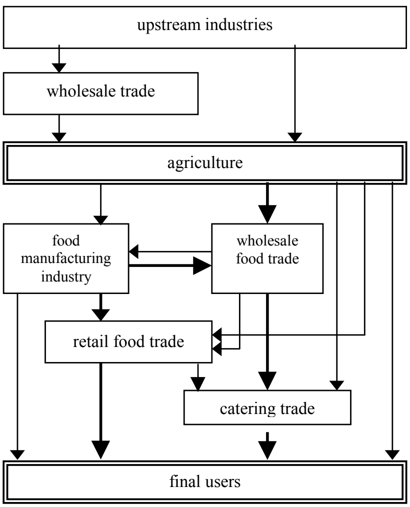
This is equally true for the downstream sector. In general, agricultural products are processed by the food industry. In addition, however, they are also sold in technical and industrial fields or used as so-called renewable resources. Thus, those sectors or sub-sectors processing agricultural produce in non-food areas could be included under agri-business, too.

Within the framework of this paper the term "agri-business" is defined as agriculture including its specific upstream and downstream economic sectors. In this context, upstream and downstream economic sectors also include those sectors frequently not engaged in profit-oriented activities, that is, economic sectors like research and development, agricultural health or accident insurances, and agricultural lobby groups. With regard to national economics these areas are adding value, so it seems to be justified to include them in the agri-business sector to determine their importance to the national economy. However, the shares of the upstream sectors whose products are used in agriculture but which were not specifically produced for the agricultural sector are not included.

The definition of the economic sectors is based on the classification of economic sectors, 1993 edition (WZ 93). Annex 1 includes a complete list of the economic sectors that form part of the agri-business sector according to the above-mentioned definition as well as their classification number under WZ 93.

Within the framework of this paper "agri-business" does not include forestry and fisheries. Therefore areas upstream of forestry with sectors manufacturing specific equipment and machinery for forestry as well as areas downstream of forestry, i.e. ranging from wood-processing industries to furniture industries, are not included, either. In Germany the forestry sector is only partly incorporated into the agricultural sector. A major part of the forestry sector consists of state forests or is managed by public bodies and other organisations, so that it seems justified to exclude it.
The upstream sector of fisheries is excluded, too. Within the downstream sectors it is more difficult to distinguish it clearly from other areas within the food industry. In principle, the processing of fish and fish products would have to be excluded from the food industry since the definition applies to terrestrial agriculture only. This has not been done, as this would, on the one hand, make calculations more complicated, particularly since it is difficult to establish an unambiguous definition. On the other hand, the calculations do not take account of import shares for the food industry. The share of the fishing industry in agri-business is quite small since Germany is a net importer of fish and fish products.
The following figure illustrates this paper's interpretation of the interdependence between agriculture and its upstream and downstream sectors. However, it must be added that the term "agri-business" has been specified even more. It is based upon the assumption that the export share of the upstream sectors has not been directly produced for the agricultural field and is thus relevant to the sectors, but not to agriculture's contribution within the framework of agri-business. Correspondingly, the importation of goods should be deducted from the downstream sectors, as imported goods add value to a national economy irrespective of the domestic agricultural sector.

Figure: *Interlockings in agri-business*



No matter whether an economic sector forms part of agri-business or not, there is always the issue of data availability, so that in the end, when it comes to a quantitative assessment, further analyses do not include all areas listed in annex 1. The calculations only take into account the economic sectors not printed in italics. As regards the sectors printed in italics, there were either no data available or they are of minor importance and thus negligible.

3. Indicators and their determination in national accounts

National accounts aim at analysing the development, use and distribution of a national economy's value added. In this context, production value and gross value added represent important indicators on the development side. The production value is based on the sales activities of a sector or industry, reflecting their economic activities. The gross value added shows the contributions individual sectors make to the gross domestic product, i.e. to the value of the output produced in Germany. The gross domestic product consists of the sum of the gross value added by the institutional sectors (plus tax on products and minus subsidies on products). Moreover, also the number of gainfully employed persons is of importance to national accounts.

In the EU national accounts are compiled according to uniform principles. In this context, Council Regulation (EC) No 2223/96 of 26 June 1996 on the European System of national and regional accounts in the Community lays down binding definitions, classifications and accounting rules prescribed in the EU. They have been developed on the basis of the United Nations' "System of National Accounts" (SNA 1993).

In accordance with this EC Regulation, the following definitions apply to production value and gainfully employed persons:

The production value is the value of all goods produced in the accounting period.

Gainfully employed persons are all persons – employed and self-employed – who engage in production activities within the production limits of the European System of national accounts.

The EC Regulation does not explicitly define the gross value added. It is deduced from the production value by deducting intermediate consumption. The gross value added reflects the value created by all units engaged in a production activity and encompasses depreciations.

In national accounts the production value is determined subject to the respective sector, whereas EC Regulation 2223/96 defines sectors as a group of institutional units with similar economic activities. An institutional unit is defined as an economic decision-maker such as joint-stock companies, cooperatives, non-profit organisations, public corporations or private households. An institutional unit can only form part of one sector. There are five groups of sectors; the non-financial joint-stock companies, the financial joint-stock companies, the state, private households and the private non-profit organisations.

The production value of non-financial joint-stock companies (i.e. enterprises) results from the value of the sales of their own production of goods and services as well as of commodities of other (domestic and foreign) economic units, plus the value of changes in their stocks of semi-finished and finished products of their own production, plus the value of fixed assets produced on own account.

The production value of the state and private non-profit organisations is determined by adding the expense items of these units. This is of importance, for instance, when it comes to the calculation of the production value of research centres.
Under the EC Regulation intermediate consumption, necessary to calculate the gross value added, measures the value of goods and services used, processed or converted in the production process. This does not include the utilisation of fixed assets, which is measured with the aid of depreciations.

4. Methods to calculate indicators of agri-business

4.1 Calculating the production value for agri-business

The production value of the agri-business sector cannot be taken from the official statistics, as they do not list an economic sector with this definition. As regards agri-business, official statistics only list the food industry's production value (including tobacco processing) for the economic sectors stated in annex 1. Thus, it is necessary to determine the production value of the other economic units of the agri-business sector by using other methods. In doing so, priority has been given to a pragmatic way of getting reliable, detailed and exact data, limiting the required work as regards data collection and processing as far as possible.
In accordance with the information given above (chapter 3), the production value of enterprises results from the value of sales of goods and services, that is from an economic unit's turnover achieved on the market. In addition, there is the value of changes in stocks and fixed assets produced on own account. Official statistics list the turnover of most of the sectors that form part of agri-business. However, data relating to changes in stocks and fixed assets produced on own account were not available and could therefore not be taken into account. This is why the sum of turnovers of the respective sectors is usually used as an estimated figure for the production value. It should come rather close to the actual production value.
However, there are some particularities in the individual economic sectors, which will be described in the following paragraphs:

a. Production value of the food industry
In the classification according to WZ 93, sub-division DA (food industry and tobacco processing) includes the classifications 15.1 to 15.9 and 16.0 (cf. annex 1). In order to calculate production value and gross value added, the production of foodstuffs for farm animals (classification 15.71) has been deducted, as they are attributed to the upstream sector.

b. Reference indicator for sales figures of the processing industry
In Germany, statistics for the processing industry are listed on the one hand for local units and for *activity parts of local units* on the other. **Local units** which produce goods for several economic sectors, the so-called combined local units, are attributed, with their total production, to those economic sectors which are of prime importance to them with regard to their number of staff. Thus, local units shifting their activities are attributed to the respective new economic sectors. For the aggregation of ***activity parts of local units,*** the results of local units are broken down into the various economic

sectors the individual operating units have to be attributed to in accordance with their type of production. This results in an exact presentation of their respective economic activities. For **enterprises** results are shown in accordance with their main economic activity, that is in line with the approach taken for local units.

In order to be able to show the agri-businesses' economic activity related to the agricultural field, statistical data – if available – are attributed to the *activity parts of local units*. This means that statistical data are used very restrictively, as thus calculations do not include other possible business areas which are not directly linked with agriculture.

Usually the *activity part of local units'* sales figures are only listed for local units with at least 20 employees. For micro-enterprises with fewer employees there are special, simplified surveys which are published in statistics regarding "industrial micro-enterprises".

However, for micro-enterprises one must assume that the total agricultural turnover is achieved domestically, as, due to the lack of data, a differentiation is not possible. One can, however, depart from the assumption that micro-enterprises, because of their size, are not engaged in exports. Moreover, it has to be taken into account that statistics break down industrial micro-enterprises into "local units" rather than into "*activity parts of local units*". Yet the difference between the two survey characteristics will probably be of minor importance, as their size will hardly result in any significant combined business activities.

c. Agricultural reference in the sub-classifications under NACE

The breakdown of economic sectors[2] helps classify the available statistical data. In the EU, the lowest level consists of four-digit classifications that are partly subdivided into five-digit classifications at national level.

This means that in the four-digit classification, for instance, the manufacture of tools can be found in group 28.62. In Germany, however, this group has been subdivided and includes, *inter alia*, the five-digit subdivision "manufacture of agricultural equipment (28.62.5)". With regard to agri-business the four-digit group's degree of aggregation is too high and does not show agriculture's share.

Official statistics only list the turnover of the four-digit group and do not show the five-digit subdivision's turnover. Thus, the turnover relating to the manufacture of agricultural equipment (28.62.5) has to be determined differently, using the value of production by the type of goods. In this context, the values of all economic goods produced by the economic sector 28.62.5 have been added, followed by a calculation of the share of this sum in the total production of goods of the economic sector 28.62. This share represents the agriculture-related share which can be attributed to agri-business. This proportional calculation, however, only takes into account the goods produced by enterprises with a minimum number of 20 employees. Yet it can be assumed that the shares relating to smaller units or smaller craft trades units hardly differ.

[2] Classification of economic sectors, 1993 edition (WZ 93), in accordance with NACE Rev.1 pursuant to Council Regulation (EEC) No 3037/90 of 9 October 1990 and Commission Regulation (EEC) No 761/93 of 24 March 1993.

d. Determining the share of agricultural buildings

Official statistics differentiate the construction industry's activities not by the target groups for which buildings are built, but group them by type of construction, that is structural and civil engineering, roofing, road construction etc. Thus, the share of agriculture-related buildings in the production value, as e.g. animal housing facilities and barns, cannot be determined on the basis of the sector's turnover.

This is why calculations have been based on data relating to gross fixed capital formation listed by economic groups. Gross fixed capital formation includes the acquisition (minus transfer) of residential buildings, non-residential buildings, equipment, farm animals and crops (afforestation, vineyards, fruit plantations etc.), intangible capital goods (e.g. copyrights, computer programmes) as well as considerable land improvements as well as land transactions costs.

Statistics show gross fixed capital formation for new installations, new equipment and other installations as well as for new buildings. In this context, new installations encompass the sum of the gross fixed capital formation for new equipment and other installations and new buildings. The calculations do not include investments in equipment, since they mainly relate to machines, machine installations, vehicles and other tangible fixed assets. Thus, calculations regarding agri-business only include gross investment expenditure for new buildings.

Therefore, the agricultural share results from agriculture's share in the total fixed capital formation in new buildings. This relative share provides the coefficient which is multiplied by production value, gross value added and the total number of gainfully employed persons in the construction industry, thus providing the respective agriculture-related values.

e. Agricultural insurances

The calculations include insurance data, excluding social insurances, in order to be able to obtain information on agricultural non-life insurances. The data are calculated with the aid of the total production value of the insurances multiplied by the agricultural share. The agricultural share results from the ratio between the insurances' input into the production of agricultural produce and the input into all ranges of production. The figures to determine the ratio were taken from the input-output tables. However, these tables are published at large intervals only, so that the calculation of the agricultural share has been based on the assumption that there have been no significant changes in the ratio determined for 1993, the last year with published data available.

f. Agricultural tax consultancy

In order to determine the share of tax consultancy flowing into agriculture, it has been assumed that the advisory services correspond to agriculture's share in the national economy. This share was then multiplied by the turnover of the overall tax consultancy sector to obtain the production value for agricultural tax consultancy. The same agricultural coefficient has been used to determine the number of gainfully employed persons and the gross value added.

It was not possible to calculate the production value the same way agricultural insurances have been calculated, since tax consultancy is not explicitly listed in the input-output tables.

g. Production value of agricultural research

It has already been mentioned that the production value of agricultural research is determined with the aid of the expense items, using the public sector's expenditure for its scientific institutions. The respective figures include universities as well as federal research centres and relate to agricultural, forestry and nutrition sciences as well as veterinary matters.

h. Taking account of export and import shares

In chapter 2, defining various terms used in this paper, it was pointed out that the export share has been deducted from the upstream sectors by making a proportional calculation of its overall turnover with regard to domestic sales, which are listed in the respective statistics.

Correspondingly, the import share should be deducted from the downstream economic sectors. This was not possible as far as the food industry is concerned (including tobacco processing). However, for food wholesale trade the import share has been deducted from the turnover. The respective import and export shares in the turnover were also used as coefficients to calculate gross value added and the number of gainfully employed persons to arrive at the share which relates to domestic activities only.

4.2 Calculating the gross value added for agri-business

The calculation of the gross value added for agri-business, too, aimed at reducing the necessary calculation work on the one hand and at obtaining very exact data on the other. Both aims are subject to data availability – such as the calculation of the production value, too. In general the remarks made about the calculation of the production value also apply to the calculation of the gross value added, as usually the same statistical sources were used and agriculture-related coefficients, which helped determine the production value, were taken for the calculation of the gross value added.

Correspondingly, the gross value added for the food industry (including tobacco processing) and agriculture (including hunting) was taken from publications of official statistics, as described above in the chapter on production value. Again, the gross value added of the sub-heading "production of foodstuffs for farm animals" has been deducted.

Moreover, the following particularities have to be pointed out as regards the calculation of the gross value added:

a. Agricultural share by using input-output tables
b. Share of gross value added in enterprises' production value applied to the production value of *activity parts of local units*
c. Calculation of the gross value-added share with the aid of data relating to cost structures

a. Agricultural share by using input-output tables

The input-output tables give an insight into the flow of goods and production interlinkages of a national economy. This information has been used in the calculation of the agricultural share of insurances to determine the share of insurance contracts concluded for agricultural activities compared with all insurance contracts that have

been concluded. This share results from the ratio between the output of all non-life insurance companies and this sector's input into agriculture. Then this share is multiplied by the gross value added of the insurance enterprises.
Since input-output tables are published at irregular intervals only, because of the extensive calculations that they require, their data are not annually available.

b. Share of gross value added in enterprises' production value applied to the production value of *activity parts of local units*

There are production value and gross value added figures of enterprises belonging to the agricultural industry. However, there are no such figures for *activity parts of local units*. Thus, national economy indicators have been calculated for the relevant *activity parts of local units*, as they enable us to achieve a more exact delimitation of the branches of business relating to agri-business. This calculation has been based on the relative share of the gross value added in enterprises' production value. This relative share has been multiplied by the production value of the *activity parts of local units* of the agro-industrial sectors, as calculated under section 4.1.

c. Calculation of the gross value-added share with the aid of data relating to cost structures

Data relating to cost structures of a specific sector, if available, enable us to deduct the share of intermediate consumption in revenues. The remaining share corresponds to the gross value-added share in the production value of the respective *activity parts of local units*. A multiplication by the production value results in the gross value added of the respective branches of business.

4.3 Determining the number of persons gainfully employed in agri-business

Gainfully employed persons are defined as all persons having one or more jobs, irrespective of the weekly hours they actually work or their weekly working hours stipulated by contract. For the definition of gainfully employed persons it does not matter whether their employment constitutes the main source of income or not. If a person has several jobs he or she is counted only once, and these persons are grouped with the economic sector in which they spend most of their working time. The data constitute annual average figures.
As in the previous sections, publications of official statistics also provided data relating to food industry (including tobacco processing) and agriculture (including hunting).
The calculations relating to the determination of the production value of agri-business, made to arrive at the agricultural share, and the relative values resulting from this, were also applied to the number of gainfully employed persons, so that the method of calculation does not differ from the above-mentioned methods.

5. Results

The total 1998 production value calculated for agri-business in Germany amounted to almost DM 1,000 billion, representing approximately 15.0 % of the overall production value. In 1998 the gross value added by agri-business amounted to about DM 220 billion, reaching a share of more than 6 % in the national economy's value added. The percentage of persons gainfully employed in agri-business amounted to almost 11 %.

Usually growth rates regarding production value and gross value added in agri-business are lower than in the overall national economy. This results in a decrease in the relative share of agri-business in the national economy indicators over time as well as in a decrease in agriculture's relative share in the national economy (cf. table).

Table : *Economic importance of agriculture and agri-business in Germany*

Sector	production value			gross value added			gainfully employed persons		
	1996	1998	2000*	1996	1998	2000*	1996	1998	2000*
	billion DM						1,000		
Upstream industries	97.6	96.0	98.6	16.0	15.9	16,1	125.2	131.1	128,1
Agriculture[1]	84.4	84.1	86.0	41.8	40.9	41.6	955.0	948.0	887.0
Downstream industries	802.7	798.4	820.6	161.4	163.5	167.4	3,082	2,952	2,938
Agri-business total	984.7	978.5	1,005.3	219.2	220.4	225.2	4,162	4,031	3,953
Overall economy	6,155.0	6,541.5	6,908.4	3,360.4	3,547.5	3,702.9	37,270	37,540	37,942
	%								
Share of agriculture[2]	1.4	1.3	1.2	1.2	1.1	1.1	2.6	2.5	2.3
Share of agri-business[2]	16.0	15.0	14.5	6.5	6.2	6.1	11.2	10.7	10.4
Share of agriculture in agri-business	8.6	8.6	8.6	19.1	18.6	18.5	22.9	23.5	22.4

*) Partially estimated – 1) Including commercial hunting, excluding fishery and forestry – 2) In the overall economy

The share of agri-business in the production value is considerably higher than its share in the gross value added. This is mainly due to trade which shows a high degree of intermediate consumption and, vice versa, a low degree of own value added. However, the trade sector contributes more than 50 % of the production value to agri-business, which is reflected in the shares of the agri-business sector. In the agricultural sector itself, however, the share in the national economy's gross value added is only slightly lower than its share in the production value.

In 1998 agri-business employed a total workforce of around 4 million people. This means that almost **every ninth job** in Germany is directly linked with agriculture. This share has been declining over time, as, on the one hand, agriculture's workforce is decreasing continuously and, on the other hand, effects of rationalisation and saturated markets in the downstream sectors result in a clear change in the number of people employed in agri-business.

The trade sector employs about 1.2 million people and thus represents the most important employer in agri-business, followed by the food industry. More than 900,000 gainfully employed persons worked for the catering trade (excluding accommodation etc.). The percentage of persons gainfully employed in agriculture (including only those persons that predominantly work in agriculture) amounted to 2.5 % of the total work-force of the national economy. Within the agri-business sector about one fifth of the workforce work in the agricultural sector itself. This means that for every job in agriculture there are four more jobs in other fields of agri-business.

6. Conclusions

This paper has tried to present methods to calculate the importance of agri-business for a national economy. The approaches taken to calculate the relevant economic indicators of gross value added, production value and workforce depend on the general availability of data. In the paper, we showed ways to determine relevant data for many branches of business attributed to agri-business to arrive at a reliable overall picture of the importance of agri-business for a national economy.

On the whole it becomes clear that the importance of the agri-business sector – as well as the importance of agriculture – is declining in spite of the growing clout of downstream sectors within agri-business. This has been shown in several other studies (e.g. Holt and Pryor (1999)) as an element of economic development. The increasing level of prosperity and a more diversified national economy result in the development of new sectors and in changes in the importance of individual sectors.

With regard to the future, however, one should keep in mind that agri-business is a sector that encompasses important new technologies (genetic engineering, bio-technology). At present, it is not possible to quantify their contribution to the gross value added. This could change in the next few years. Furthermore, renewable resources and goods and services formerly called "non-marketable" (like environmental amenities from agriculture) will gain importance. Then, there will probably be a new assessment of the importance of agriculture and agri-business.

References

Holt, T., Pryor, S. (1999), Agribusiness as an Engine of Growth in Developing Countries. USDA ERS paper.

Annex: List of economic sectors that form part of agri-business

Section of WZ 93	short name
Upstream industries:	
14.30	Mining of chemical and fertilizer minerals
15.71	Manufacture of prepared feeds for farm animals
22.13.1	*Publication of technical journals (technical media)*
24.15	Manufacture of fertilizers and nitrogen compounds
24.20	Manufacture of pesticides and other agro-chemical products
28.62.5	Manufacture of tools for agriculture
29.3	Manufacture of agricultural and forestry machinery
ex 45	Construction (agricultural construction)
45.11.3	*Land improvement and renaturation of water bodies*
51.11	Agents involved in the sale of agricultural raw materials, live animals, textile raw materials and semi-finished goods
51.2	Wholesale of agricultural raw materials and live animals
51.55.4	Wholesale of chemical-technical products
51.55.5	wholesale of fertilizers
51.66	Wholesale of agricultural machinery and accessories and implements, including tractors
51.7	Other wholesale
63.12	*Storage and warehousing*
65.12.7	*Credit institutions fulfilling special tasks (farm mortgage bank)*

66.03.2	Non-life insurance (agricultural insurance)
71.31.0	*Renting agricultural machinery and equipment*
73.10.3	*Research and development in the areas of agricultural, forestry, and nutritional research*
74.12.3	Tax consultancy (agric.)
74.14.1	*Business and management consultancy*
74.40.2	*Launching and conducting advertising campaigns (agricultural advertising)*
74.84.2	*Experts (expertise)*
75.30.3	*Old age security for farmers*
75.30.5	*Statutory health insurance (agricultural health insurance fund)*
75.30.7	*Statutory accident insurance (agricultural employers' liability insurance association)*
85.20	*Veterinary activities*
91.11.1	Activities of economic associations (advocacy)
91.11.2	*Activities of economic associations under public law (chambers of agriculture)*
Agriculture:	
01.1	Growing of crops; market gardening; horticulture
01.2	Farming of animals
01.3	Growing of crops combined with farming of animals (mixed farming)
01.4	Agricultural and animal husbandry service activities, except veterinary activities
01.5	Hunting, trapping and game propagation, including related service activities
Downstream industries:	
15.1	Production, processing and preserving of meat and meat products
15.2	Processing and preserving of fish and fish products
15.3	Processing and preserving of fruit and vegetables
15.4	Manufacture of vegetable and animal oils and fats
15.5	Manufacture of dairy products
15.6	Manufacture of grain mill products, starches and starch products
15.7	Manufacture of prepared animal feeds
15.8	Manufacture of other food products (excl. manuf. of beverages)
15.9	Manufacture of beverages
16.0	Manufacture of tobacco products
51.17	Agents involved in the sale of food, beverages and tobacco
51.3	Wholesale of food, beverages and tobacco
52.11	Retail sale in non-specialized stores with food, beverages or tobacco predominating
52.2	Retail sale of food, beverages and tobacco in specialized stores
52.61	Retail sale via mail order houses
52.62	Retail sale via stalls and markets
ex 55	Catering trade

6 June 2001

Invited Paper Session

AGRICULTURAL CENSUS AND SURVEY ISSUES

Chair: F. Hardy

The Agriculture Census - Purpose, Characteristics and Best Practices

Séghir Bouzaffour
Direction de la Programmation et des Affairs Économiques, Ministère de l'Agriculture, Avenue Hassan II, Km 4 - Station Debagh 1, Rabat, Maroc,
e-mail: bouzaffour@dpae.madrpm.gov.ma

Michael Trant
Agriculture Division, Statistics Canada, 12-D8 Jean Talon Building, Ottawa, Ontario K1A 0T6, Canada, e-mail: mike.trant@statcan.ca

Abstract: A census is a valuable statistical tool that deserves consideration. Too often the arguments for a census are dismissed early in the planning process for cost reasons despite the need for geographic and commodity detail, a list of agricultural holdings and benchmark data. The strengths of a census lie in the collection of data from all members of the population. When well managed, a census can provide the coverage and detail that are not possible from sample surveys. Translating the characteristics of a good census into a successful census program can be a challenge but there are many good examples so the task is well within the grasp of most countries.

Keywords: census, characteristics, best practices

1. Introduction

In many countries the Census of Agriculture remains the main source of agricultural statistics and the base for the farm register, the intercensal sample surveys and related statistical activities. This paper reviews the purpose and characteristics of a census of agriculture and lists a number of examples of "best practices" that have been successfully used in both developed and developing countries. The paper is not a guide on how to conduct a census. That information is available from the Food and Agriculture Organisation publication, "Conducting Agricultural Censuses and Surveys" (FAO, 1995).

Agricultural data collection programs can normally be classified as either census or sample surveys. A census is a survey, which collects data from all members of a population. A sample survey differs by surveying a small number or proportion of the population. Agriculture statistics programs benefit from being able to draw on information from census surveys, sample surveys and administrative sources of information such as customs documents. Each has a contribution to make to the information system. Government agencies collected most of their data using a census until the early 1940's. The extensive use of sample surveying in place of a census is a relatively recent development.

2. Purpose and Objectives of an Agriculture Census

Despite the advantages of sample surveys, and the increasing amount of timely and reliable administrative information, there is still an important role for a census in the statistical information system. In principle, a census is able to provide estimates for each agricultural holding and can provide information that can not often be supplied by administrative data or sample surveys. The classic arguments for conducting a census are:

(1) establish a list of all known agricultural holdings for sample surveying,
(2) establish benchmark reference levels for the annual survey program.
(3) provide detail for small geographic areas, and
(4) provide information for commodities where production is limited.

A census of agriculture provides an image of a nation's farming and agricultural activity at a particular point in time, with a single reference date for the information. Working from a comprehensive list of farms and farmers or using maps to locate and visit farms, enumerators attempt to interview or leave a census of agriculture questionnaire at every farm holding or with every farm operator. Information is collected from all holdings regardless of size or geographic location.

3. Characteristics

The characteristics of a good census of agriculture are similar to those of any survey program. The overall objective is to provide not only quality data but also ensure that one is measuring the right things. If the concepts and the things being measured are not reasonably accurate reflections of the real world, then no amount of sophisticated statistical technique or funding will produce useful numbers (Bonnen, 1975).

The strengths of a census lie in the collection of data from all members of the population. When well managed, a census can provide the coverage and detail by geographic region and the size and type of farm that cannot be obtained otherwise. The census is an archive of this benchmark information, furnishing very valuable base line information useful for historical comparisons.

A census has limitations, the most significant being the cost and the logistics associated with the need for substantial human and financial resources for a relatively short period of time. As a consequence, it has never been very practical to conduct a census on an annual basis. With a large number of agricultural holdings, complete enumeration is an expensive and time-consuming operation. The size of the census operation is a significant challenge. A large number of people and resources need to be organised and mobilised to collect and process a large volume of information in a short period of time.

There are no sampling errors associated with a census but a census is subject to non-sampling errors. Non-sampling errors arise during the course of virtually all survey operations and become awkward and complex to manage and control as sample size increases. The errors of concern include errors in the frame, difficulties in establishing precise operational definitions, interpretation of the questions by respondents, inconsistencies in field collection procedures or mistakes in the processing operations.

A census of agriculture is an attempt to enumerate all farms but a census is also almost always incomplete to some extent. It is inevitable that there will be some undercoverage of farms. When conducted effectively, however, a census can provide a comprehensive listing of farm operations for subsequent sample surveys and a level of geographic and commodity detail that is not possible with sample surveys.

An agriculture census is not a practical means of providing annual or seasonal agricultural data on crop or livestock production. Sample surveys are more efficient and are able to provide results in a more timely fashion. Even if the resources to collect the data using a census were available, the large quantity of data to be collected and processed usually means that the results cannot be released in less than a year following the data collection period. This is the case in Canada and the United States. In countries where resources are more limited and the collection and processing must be extended, results will take even longer to be released.

4. Best Practices

Translating the characteristics of a good census into a successful census program can be a challenge but there are many good examples of census of agriculture programs so the task is well within the grasp of most countries. The objective of this section is to review some of the processes involved in planning and implementing a census, identifying whenever possible examples of best practices pioneered by others. It is not feasible to list all best practices. As a consequence, the paper is limited to those the authors believe to be the most generally applicable. The examples draw heavily on the experiences of Canada and Morocco, countries the authors know best.

4.1 Project Planning

It is critical to have a clear understanding and to formulate a clear statement of the objectives in terms of what is to be measured and why, the data quality expected, budget constraints and expected delivery dates. Stating the objectives is the first step in forming a consensus among the data users and the statisticians and facilitates decisions with respect to the project's design.

Canada has a policy to develop census objectives in partnership with major data users and other interested parties. Before there is any major design or redesign there is extensive user consultation to identify the content, options and develop public support for the program well before it reaches the data collection stage. Where explicit data quality targets exist, they are written into the census objectives. The purpose of the Census of Agriculture is also reviewed periodically to ensure that it remains relevant to users whose needs may be evolving and changing.

The approach in Morocco is very similar. In Morocco, as in almost all developing countries, a census of agriculture, enumerating all farmers is not a common occurrence and the census initiative requires a special effort. Since the 1960's, the statistical system has been the subject of regular reviews every 2 to 3 years. In conjunction with these reviews, statistical meetings are held to examine survey results and the methodology used so that needed improvements can be made in methods, concepts and definitions. The current agriculture statistics system is the result of a 12-year economic and statistics development project between 1983 and 1994, supported with technical assistance from the United States Department of Agriculture.

4.2 Concepts, Definitions and Classifications

In order to use data, one must understand the concepts and definitions underlying the information and understand what the data purport to measure. Although the use of international standards for concepts, definitions and classifications will assist the survey developers, what is eventually put in place may differ from international conventions in order to address local conditions or meet user requirements. Concepts, definitions and classifications need to be documented and any differences from international standards or normal conventions should be noted.

Both Canada and Morocco publish standard classifications of industries, products, occupations and geography. Concordances between concepts, definitions and classifications and international standards produced by the United Nations Statistical Office, the International Labour Office, Organisation for Economic Co-operation and Development and Eurostat are maintained as are those between different vintages of the standards to assist with comparisons.

4.3 Questionnaires

Good questionnaire design plays a central role in the data collection process. The questionnaire has a major impact on data quality, respondent burden, interviewer performance and respondent relations. The design of questionnaires takes into account the statistical requirements of data users, administrative requirements of the survey organisation and the requirements for data processing, as well as the nature and characteristics of the respondent population. Good questionnaires can both minimise respondent burden and collect the required information with a minimum of errors.

Some of the best practices are also sometimes the most obvious and easiest to implement. In both Canada and Morocco efforts are made to use words and concepts that have the same meanings for both respondents and statisticians. This may appear obvious but it requires a lot of work and careful testing to achieve. To the extent possible, concepts and wording are harmonized with those already in use and there is no hesitation to reuse successful questions or techniques borrowed from other surveys. Questionnaires also need to be designed to be attractive and easy to complete. Questionnaires are always pretested with an informal test that helps identify the most obvious problems such as poor wording, ordering of questions, layout or instructions. Finally, a pilot test is always conducted following the pretest to observe how all the census operations work together.

4.4 Data Collection Operations

The individuals and organisations that complete questionnaires without payment are a survey taking organisation's most valuable resource. The impact of data collection operations on data quality is critical, as the data are the primary input to the final estimates. Sophisticated statistical procedures are of little help in correcting the effects of flawed collection operations. Interviewers are an essential element of the collection operation so it is critical that the interviewers are given the appropriate training and tools to ensure success.

In both Canada and Morocco, efforts to maintain an efficient and effective data collection operation centre around careful planning, testing, quality control and interviewer training. Interviewing, manual editing, coding and data entry are some of the complex activities done by census employees and the use of quality control procedures are particularly useful in providing feedback reports of information on frequencies of and causes of error for the census manager, operations employees and statisticians.

4.5 Response

Despite the best efforts of survey staff to maximise response, some non-response is certain to occur. Farmers are required by law to respond to the Census of Agriculture in most countries, but compliance is dependent on the co-operation of respondents and their willingness to report to the best of their ability. There are always some respondents who do not answer parts or all of a questionnaire and some, for good reasons, are not or cannot be contacted. The degree to which non-response is pursued is based on budget and time constraints or the impact of non-response on the estimates. Non-response is important to minimise as it can contribute to bias in the final estimates when the characteristics of non-respondents differ from those of respondents.

Substantial effort is made in both Canada and Morocco to minimise non-response. Whenever possible, efforts are made to follow-up the non-respondents. The non-respondent follow-up increases the response rate, highlights shortfalls in the data collection procedures and can help identify to some extent, whether respondents and non-respondents have similar characteristics.

Cases of non-response are unusual in Morocco where local administrative authorities provide assistance in the completion of questionnaires. Surveying farmers in close co-operation with local Moroccan authorities ensures good response and participation from the farming community for all survey and census activities regardless of the data sought. In Morocco the non-response rate and number of refusals is very low; low enough in any case, to preclude the need for any specific corrective measures. For example, of the 1.5 million farms, the total number of refusals or failures to respond did not exceed .2% or about 2 per thousand. In Canada the comparative figure is about 2% or 20 per thousand.

4.6 Editing

Data editing is the application of a review process that identifies missing, invalid or inconsistent entries or points to potential errors in data records. Editing encompasses a wide variety of activities, ranging from interviewer field checks, computer generated warnings at the time of data capture, identification of units for follow-up, through to complex relationship verifications, error localization and data validation.

The goals of editing are, (1) to provide the basis for future improvement of the survey vehicle, (2) to provide information about the quality of the survey data, and (3) to tidy-up the data (Granquist, 1984). Most statistical organisations spend a disproportionate amount of resources on "tidying-up the data". While it is important to remove invalid or inconsistent entries to maintain the statistical agency's credibility and reputation and to facilitate further data processing and analysis, statisticians should resist the temptation to over-edit.

Data editing is believed to be the single most expensive activity in a census or survey and editing has been estimated to account for as much as 40% of the total survey budget in the case of business surveys (Gagnon, Gaugh and Yeo, 1994). The practice of over editing is costly in terms of resources, timeliness and respondent burden, all of which can lead to severe biases from fitting data to implicit models imposed by the edits.

Canada and Morocco "macro edit" individual records with a focus on inconsistencies that are apparent, not only in the individual records, but also in the published estimates.

Editing can serve a useful purpose in cleaning-up data files but it serves a more useful role in providing information about the survey process. Editing is invaluable in improving definitions, the questionnaire, interviewer instructions and field procedures. A system to monitor the edit process, produce audit trails, diagnostics and performance measures is essential for making best use of the edits and improving census procedures.

4.7 Imputation

Imputation is the process used to resolve problems associated with missing, invalid or inconsistent data identified during editing. Imputation changes some of the data values to ensure that a believable and internally coherent record is created. Imputation may be automated, manual or a bit of both. Good imputation systems limit the bias caused by not having observed all the desired values. They also provide an audit trail for evaluation purposes and ensure that imputed records are internally consistent.

Imputation methods can be classified as either stochastic or deterministic, depending on whether there is randomness in the imputed data (Kalton and Kasprzyk, 1986, Kovar and Whitridge, 1995). Deterministic imputation methods include logical, historical, mean, ratio and regression methods and the nearest-neighbour approach. Stochastic imputation methods include the hot deck, regression with random residuals and any other deterministic method with random residuals.

Canada and Morocco both use imputation methods that are automated, objective and replicable. It is not always necessary to develop a new imputation system for each application. Existing generalised systems are available from a number of statistical agencies, all readily adaptable to various survey and census situations. Labeling the imputed values and clearly identifying the methods and sources is however necessary. The Canadian Census of Agriculture database retains the imputed and unimputed values of each variable for each record.

Morocco uses a system quite similar to that adopted by Canada. In all surveys including the Census, the imputation system involves the use of a proven computerised program. The database, containing the data from the various surveys, the Census program and any other sector studies, serves as the historical archive for all the information processed and compiled by the Ministry of Agriculture.

4.8 Estimation

Estimation is the task of assigning values to unknown population parameters by tabulating information from a data set. Examples include simple descriptive statistics such as totals, means, and ratios. Census estimates are dependent on the quality of the information collected and the impact of the data processing procedures specifically, data capture, edits and imputations. Auxiliary data can also make a major contribution to the quality of the final census estimates. Whenever auxiliary data, such as those that are available from administrative data systems, are accessible, they should be compared to the census estimates and efforts made to reconcile or explain the differences. This will normally improve precision and leads to greater consistency between estimates from different sources, an important issue to data users.

In both Canada and Morocco the quality of the estimates is largely dependent on the quality of the information collected from respondents and the impact of the data capture, edit and imputation procedures. Census estimation is essentially an exercise in computation where automation and efficiency are important. Computing the census estimates is fundamentally a mechanical operation.

4.9 Data Quality Evaluation

Evaluation of data quality is the process of reviewing the final estimates with respect to the accuracy and reliability of the data. There are generally two reasons for data quality evaluations. First, the process of reviewing the data prior to the official release ensures that grossly erroneous data are not released and that data of marginal quality are identified. Second, the evaluation provides quantitative information on specific sources of error in the data. Statistical agencies have a responsibility to conduct data quality evaluations and inform data users of the results of the evaluations. Data users are not in a position to make such assessments but most need the information on data quality to assess the degree to which their use of the data may be limited.

Data evaluations are an essential element of all statistical programs. Evaluations can be used to improve data quality on the next occasion of the census when errors can be traced to specific steps in the census process, such as the undercoverage of farms. In Morocco, evaluation is largely based on specific control surveys for monitoring the main annual surveys and on postcensal surveys for agricultural census operations. Such evaluation tools make it possible to determine the coverage rate for the target population and help identify the bias associated with procedures and processes.

Quality is not something that should be maximised at all costs. The statistician's primary objective should be to establish and maintain an appropriate balance between the relevance, timeliness and quality of the information given specific resources. The challenge is to make the appropriate trade-off among the evolving data needs, costs, response burden and the various dimensions of quality.

4.10 Disclosure Control

Disclosure control refers to the measures taken to protect respondent information from accidental disclosure. Respondents provide information to statistical agencies in confidence. Statistical agencies collect information for statistical purposes and the individual information of respondents should never be made available to others except as aggregates. Furthermore, it is the responsibility of the statistical agency to ensure that confidentiality remains uncompromised by residual disclosures resulting from multiple releases of information from a database, as can be the case with census data that are cross-tabulated. The confidentiality of individuals and businesses must be protected.

In Canada the legal provisions of the "Statistics Act" govern disclosure. The confidentiality provisions of the Act are rigorous.

> *"no person who has been sworn in under section 6 shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organisation."*

In Morocco, the law governing the confidentiality of census and survey data is equally rigorous. The law and the efforts of the government of Morocco, to protect information, is the basis of the trust and co-operation between respondents and enumerators, which makes a significant contribution to the quality of the information collected. Information provided by farm operators is never disseminated on an individual level. The information is only made available and published on an aggregate level. As is the case in Canada, the law provides legal recourse, and severe penalties to employees who fail to protect the confidentiality of information on individuals, businesses and organisations, regardless of their rank or function.

4.11 Data Dissemination

Absolutely critical to the success of any statistical program is the requirement that the statistical agency and the chief statistician be independent and free of political influence with regard to the content and timing of any statistical release (Fellegi 1995). It is important that attention be paid to the delivery of information to users to ensure that it is accurate, complete, accessible, understandable, usable, timely and appropriately priced. Data must be verified prior to release to ensure they are consistent with the source data and electronic products must be tested to ensure they perform as planned. Those responsible for dissemination need to be aware of user needs, interests and capabilities for analysing and processing statistical information. It is also important for the statistical agency to be able to provide the data in a format to suit user needs on diskette, CD-ROM, the internet, paper publications, microfiche, or by telephone, facsimile, public presentation or a television or radio interview.

To achieve such objectives in Canada efforts are made to: (1) release data to all Canadians at the same time with no exceptions at www.statcan.ca, (2) provide documentation with all statistical estimates regarding descriptions of its quality and the methodology, and (3) provide a contact person, a telephone number and address with each release and ensure that there is a prompt and knowledgeable response to all inquiries.

In Morocco, the manner in which agricultural information is disseminated differs slightly from that of Canada. The results are first announced through press briefings, radio and television. This is the responsibility of the Ministry of Agriculture. The information is then sent to a number of focal points in other ministries, training institutions and to the Ministry of Agriculture's entire central and regional network. A ministerial documentation centre then makes the documentation available to researchers or any other users, to be loaned out or for on-site consultation. Lastly, all statistical information is posted on the Ministry of Agriculture Web Site www.madrpm.gov.ma, making the information available free of charge to all data users.

4.12 Documentation

Documentation constitutes a record of the statistical activity, including the underlying concepts, definitions and methods used in the production of the data. Documentation may consist of a series of reports serving quite separate needs. There may be a report for management, technical staff, planners of other censuses, and data users. Documentation contains, the objectives of the statistical program, the questionnaire, concepts and definitions, information on any field and pilot tests, methodology, information on the processing systems, manuals and operational procedures, quality control, data quality, resources and budget.

5. Conclusions and Observations

The Census of Agriculture remains the main source of agricultural statistics and the base for the farm register, intercensal sample surveys and related statistical activities in many countries. Agriculture statistics programs benefit from being able to draw on information from census surveys, sample surveys and administrative sources such as customs documents. Each has a contribution to make to the information system.

Despite the advantages of sample surveys, and the increasing amount of timely and reliable administrative information, there is still an important role for a census in the statistical information system. The Census collects data from all members of the population and provides estimates for each agricultural holding to establish a farm register, benchmark estimates, and provide geographic and commody detail not possible with sample surveys.

A census has limitations, the most significant being the cost and the logistics associated with the need for substantial human and financial resources for a relatively short period of time. A task that is often further complicated when respondents speak various languages and have a wide range in literacy skills.

The sound statistical practices applicable to sample surveys are also applicable to a census. A census is little more than a very large sample survey without "textbook" sampling error but with major challenges in terms of logistics and communications. The strengths of a census lie in its ability to provide detail, commodity detail, geographic detail, benchmark estimates, and enough records to permit cross-classifications that are just not possible from sample surveys.

A census is a valuable statistical tool that deserves consideration. Too often the arguments for a census are dismissed early in the planning process for cost reasons despite the need for geographic and commodity detail, a list of agricultural holdings and benchmark data, information that only a census can provide.

References

Bonnen, J.T. (1975) Improving Information on Agriculture and Rural Life, *American Journal of Agricultural Economics*, Vol. 57, No. 5, December 1975, p. 757.

FAO, (1995) Conducting Agricultural Censuses and Surveys, *FAO Statistical Development Series, No. 6*, Rome.

Fellegi, I.P. (1995) Characteristics of an Effective Statistical System, *Morris Hansen Lecture, Washington Statistical Society,* Washington, D.C.

Gagnon, F., Gough, H. and Yeo, D. (1994) "Survey of Editing Practices in Statistics Canada, *Statistics Canada Technical Report.*

Granquist, L. (1984) On the Role of Editing, *Statistisk tidskrift,* Vol. 2, pp. 105-108.

Kalton, G. and Kasprzyk, D. (1986) The Treatment of Missing Survey Data, *Survey Methodology, Vol. 12, pp.1-16.*

Kovar, J.G. and Whitridge, P. (1995) Imputation of Business Survey Data, *Business Survey Methods,* B.G. Cox et al. (eds.). Wiley, New York, pp. 403-423.

Statistics Canada (1998) Policy on Standards, *Policy Manual, Section, 2.10.*

Development of Agricultural Statistical Methods for the Russian Federation’s New Economy

Martha S. Farrar,
USDA, NASS, South Bldg -1400 Independence Ave SW
Washington, D.C. 20250-2013, D.C., USA
e-mail: mfarrar@nass.usda.gov

Alexandra V. Epikhina,
State Committee of the Russian Federation on Statistics (GOSKOMSTAT of Russia),
Myasnitskaya ul. 39, 103450 Moscow, Russian Federation
e-mail: epihina@gks.ru

Michael A. Steiner,
USDA, NASS, South Bldg -1400 Independence Ave SW
Washington, D.C. 20250-2013, D.C., USA
e-mail: msteiner@nass.usda.gov

Abstract: The State Committee of the Russian Federation on Statistics (GOSKOMSTAT of Russia) and the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) have collaborated since 1995 on improving statistical survey methodology for agricultural surveys in the Russian Federation. The 1990's brought major structural changes in the agricultural sector of the Russian Federation. With the development of the private sector and increase in production from private farms and households, there was a need to expand the use of surveys into a sophisticated statistical system. This paper describes the new agricultural surveys implemented in the Russian Federation, discusses the methods used in the surveys’ design and procedures, and addresses issues faced in developing and implementing these surveys.

1. The Russian Federation

The Russian Federation is the largest country on earth, spanning one-eighth of the world’s land-mass. Its territory occupies a major portion of Eastern Europe and Northern Asia with borders on the Arctic Ocean to the North, the Baltic and Azov Seas to the West, the Pacific Ocean to the East, and the Black and Caspian Seas to the South. Although the Russian Federation crosses many climatic zones, the majority of its land is in the temperate zone with 61% subject to permafrost conditions. Russia’s land area is composed of 50% forest and only 13% agricultural land. Climatic conditions throughout the Russian Federation are highly variable with extreme temperature ranges in the summer and in winter. A major portion of the land lies within areas considered as “high risk” or unfavorable to agriculture, posing challenges to agricultural production.
The Russian Federation is divided into 89 administrative and political entities referred to as regions. Included in the 89 regions are a total of 21 republics, 6 krays (federal republics), 2 federal cities (Moscow and St. Petersburg), 49 Oblasts, 1 autonomous

region and 10 autonomous areas. Russia is one of the most sparsely populated countries in the world (8.5 persons per square kilometer) with approximately 145.2 million people who are predominantly (73%) urban dwellers. Russia is a multiethnic society with over 100 ethnic groups. Four-fifths of its population is Russian.

2. Agricultural Sector Reform

Profound economic changes have taken place in the Russian Federation due to the agrarian reform programs of the early 1990's. The impact of these programs resulted in furthering the transition from a planned to a market economy as land policy changed, new entities in the production of agriculture emerged, and the private sector experienced active growth in agriculture production. The privatization of land meant the government monopoly of land property had collapsed, that agricultural land would be reallocated to include the private sector, and that a land market was being developed. Rules governing privatization varied by regions, but predominantly, 1) persons leaving the employment of state farms could be given a share of its land or 2) persons could apply to local administrations to acquire land. Persons with land could rent, trade, sell, or combine land with other individuals with property.

The Russian Federation Land Valuation Service reported that as of the year 2000, 20% of the total land was held in private ownership while 80% remained the property of federal or municipal governments, of which, 44% of the total agricultural land was officially transferred for agricultural use and 35% of the total agricultural land was held under rental agreement by agricultural producers.

With the reallocation of land, the agricultural sector of the Russian Federation now has three major types of agricultural businesses: Agricultural Enterprises, Private Farms, and Private Households. Prior to the 1990s, three-fourths of the national agricultural gross output had been produced by agricultural enterprises. With land redistribution, Russia witnessed a dramatic growth in agricultural production within in the private sector.

Agricultural Enterprises: Agricultural Enterprises include collective and state farms, joint stock companies and partnerships of all types, and other special arrangements. Currently they include 27,000 large and medium-size farming units which average 5,600 hectares of agricultural land, 2,700 hectares of cropland, 600 head of cattle, 250 swine, sheep, and goats, and 190 employees.

Private Households: Private households include rural households with land estates, collective and individual orchards and kitchen gardens. There are approximately 16 million rural households with land estates (average land area of 0.4 hectares per household). Additionally, there are 19 million families with collective and individual orchards and kitchen gardens (average land of 0.09 hectares per family).

Private Farms: Private farming is a form of free enterprise whereby producers use owned or rented lands, and operate to produce, process and market agricultural commodities. With land reform, Private Farms emerged in the agricultural sector in the 1990's. Agricultural land operated by private farms increased from 4,400 farms with 100,000 hectares in 1990 to 261,700 farms with 14.3 million hectares in 2000 (average land area of 55 hectares per private farm).

Statistical data showing total agricultural output and production of major agricultural commodities by types of agricultural businesses are given in Table 1 for 1990, 1995, and 2000.

Table 1: *Major agricultural output by types of agricultural business (% of total)*

	1990	1995	2000
Total Agricultural Output			
Agricultural enterprises	73.7	50.2	43.1
Private households	26.3	47.9	53.9
Private farms	0.0	1.9	3.0
Cereal crops (weight after cleaning)			
Agricultural enterprises	99.7	94.4	90.7
Private households	0.3	0.9	0.9
Private farms	0.01	4.7	8.4
Potatoes			
Agricultural enterprises	33.9	9.2	6.5
Private households	66.1	89.9	92.4
Private farms	0.0	0.9	1.1
Vegetables			
Agricultural enterprises	69.9	25.3	19.9
Private households	30.1	73.4	77.9
Private farms	0.0	1.3	2.2
Fruits and berries			
Agricultural enterprises	49.5	22.5	13.6
Private households	50.5	77.4	86.3
Private farms	0.0	0.1	0.1
Livestock and poultry for the slaughter market (in slaughter weight)			
Agricultural enterprises	75.2	49.9	40.7
Private households	24.8	48.6	57.4
Private farms	0.01	1.5	1.9
Milk			
Agricultural enterprises	76.2	57.1	47.8
Private households	23.8	41.4	50.5
Private farms	0.0	1.5	1.7
Eggs			
Agricultural enterprises	78.4	69.4	70.8
Private households	21.6	30.2	28.8
Private farms	0.0	0.4	0.4

3. Agricultural Statistics System of the Russian Federation

GOSKOMSTAT of Russia is responsible for the majority of agricultural data collection and analysis for the Russian Federation. The agency's major agricultural functions are to co-ordinate and manage the collection of agricultural data from the various types of operations and from other federal executing agencies; to provide methodological and informational assistance in computation of macroeconomic indicators for the National Accounts System, to develop food resource budgets, and to conduct statistical observations of agricultural businesses to monitor the trends in agriculture in the new economic environment.

The general goals and statistical procedures used in surveying social and economic changes taking place in Russia are identified by the Federal Program of National Statistics Collection and Management. This Program, which is approved annually by the government of the Russian Federation, specifies the kind of data to be collected, analyzed and reported by federal executing agencies within the range of their responsibilities. The Program ensures access to the informational resources available from different ministries and governmental agencies, provides standards for statistical procedures and principles, and promotes uniformity in statistical information throughout the federal level.

GOSKOMSTAT of Russia collects, manages and reports data for the following areas:

- Level of employment
- Level of investments
- Agricultural commodity supply, intermediate consumption, value added
- Crop production
- Livestock inventory and production
- Average prices and producer's price indexes
- Agricultural producer's financial situation
- Food resource budgets

Additionally, this Program provides for major areas of non-centralized agricultural statistical observation and specifies when data are collected and analyzed by various divisions within the Ministry of Agriculture, Ministry of Finance, Russian Federation Land Valuation Service and other federal agencies.

Table 2 below, compares the current agricultural statistics program of the Russian Federation with the program prior to 1990.

Table 2: *Agricultural Statistics by Type of Agricultural Business*

	Prior to 1990	Current
Agricultural Enterprise	Complete enumerations: *Crops*- annual, every 2 weeks during the sowing and harvest seasons *Livestock* – annual, monthly *Economic* - annual accounting document	Complete enumerations: *Crops and livestock* – annual, monthly (integrated form) *Economic** - annual accounting document
Private Farms	Did not exist	*Crops* – planted areas – annual complete enumeration *Crops* - production** – annual survey *Livestock*** – annual survey *Economic* – summer 2000 was first survey
Private Households	*Crop and livestock*: statistics based on annual surveys using sample for Family Budget Survey – (not designed for agriculture, not all regions represented) and data from local administrative offices. Census every 10-15 years	*Crop, livestock, and economic:* statistics based on multi-use monthly survey (new sample stratified for agriculture used in each region-beginning in 1998) and data from the Land Management and Evaluation committee and local administrative offices. Census every 10-15 years

* A special economic survey is being developed for 2001.

** Census for crop production and livestock for private farms is conducted every third year. Most recent census conducted January 2000 (data for 1999).

4. Collaboration With NASS

GOSKOMSTAT of Russia has collaborated with NASS of USDA since 1995 to improve the statistical survey methodology for agricultural surveys in the Russian Federation. Areas of joint work include the design and allocation of samples, questionnaire design, data collection strategies and management, procedures for statistical analysis, statistical estimation and dissemination of agricultural data.

NASS has provided numerous training opportunities in Russia and in the United States for the staff at GOSKOMSTAT of Russia on the practical aspects and application of sample survey methodology. Extensive technical assistance was provided in the area of sample design and survey methodology for two new advanced surveys: the Rural Household Farming Survey and the Private Farm Survey. Specific survey activities supported by NASS include: (1) Pilot Survey of Private Farmers (Mikhailov District of the Ryazan Oblast, 1996); (2) Pilot Survey of Rural Households (Kirov and Penza Oblasts, 1997); (3) National Rural Household Farming Survey (underway since 1998); (4) Pilot Economic Survey of Private Farms (Tver Oblast, 2000); and (5) National

Economic Survey of Private Farms (2000).
NASS provided technical assistance in the preparation of two GOSKOMSTAT of Russia statistical reports. The 1999 publication, *Private Household Farming in Russia*, and the 2000 publication, *Agricultural Activity of Private Farms in Russia*, are available in English and Russian. Funding for NASS collaboration with GOSKOMSTAT of Russia has been provided under the Emerging Markets Program administered by the USDA Foreign Agricultural Service.

5. Use of Sample Selection Procedures to Survey Agricultural Businesses

5.1 Improved data collection and management procedures

GOSKOMSTAT of Russia recognized that the "complete enumeration" mode of data collection that had traditionally been used by the national statistics system was no longer cost effective under the changing conditions in Russia. With multiple types of agricultural operations, increased importance of the private household sector, disintegration of large agricultural enterprises and establishment of private farming, GOSKOMSTAT of Russia actively pursued developing an expanded sophisticated *statistical survey system* to improve national statistics and to produce results in an efficient and timely manner.
A statistical survey system requires a sampling frame. The Agricultural Commodity Producers' sub-register of the Unified National Register of Enterprises and Organizations consists of two parts that serve as sampling frames: the Agricultural Enterprise sub-register and Private Farm sub-register. The Land Taxpayers regional list served as a sample frame for private households.
Starting in 1998, a new, national survey was initiated for the rural household farming sector. The survey was conducted monthly throughout all regions of Russia. On a national level, agricultural activity of private farms was captured in surveys, while it was not until the year 2000, under the collaboration program, that the first National Economic Survey of private farms was designed, successfully administered and completed in Russia. Additionally, special surveys of private farms and of large and medium-size agricultural enterprises were conducted that closely monitor new processes and factors emerging in the agricultural sector. The Agricultural Business Condition Survey which permits analysis of issues related to the allocation of land and property shares, dividend payments, expansion of commercial operations, improvement of credit and financial mechanisms, situation on the agricultural commodity markets has been conducted over the past few years.
The following describes the sample designs and methodological approaches used in each of the newly designed surveys: the Rural Household Farming Survey and the Private Farm Survey.

5.2 Rural Household Farming Survey

Crop and livestock statistical information on rural household farming had been collected through a periodic (10-15 years increment) census (which included data on hectares planted to agricultural crops, fruits and berries, and livestock inventory), from local committees on land resources and management and rural administrations, and through an annual sample survey. The sample for the survey was drawn for a Family Budget Survey – and was not allocated for agriculture nor were all regions represented.

The Rural Household Farming Survey program includes the description of hectares operated; planted and harvested areas; crop yields; livestock and poultry inventory; livestock production; livestock and poultry fodder consumption; household in-farm stocks of agricultural commodities; and costs of purchased food products and consumer services.

To improve data collection and management for rural household farming, a new private household survey was developed in 1998 to represent the rural population residing in all regions of the Russian Federation. The sample covers 15,000 households (0.1% of the total households) throughout the country.

Lists of land taxpayers has served as sampling frames for the Rural Household Farming Survey. These lists, which include landowners, land users and individuals holding lifelong interests and rights of succession, are maintained at district and city committees of land resources and management according to the Russian Federation Government Resolution.

A four-stage sample design was developed to ensure coverage of the target population and to ensure appropriate geographic representation.

Stage I: A list of administrative districts was compiled in serpentine order from north to south using geographical and topographical maps of administrative and political entities. The listing included the number of households and accumulated land by district.

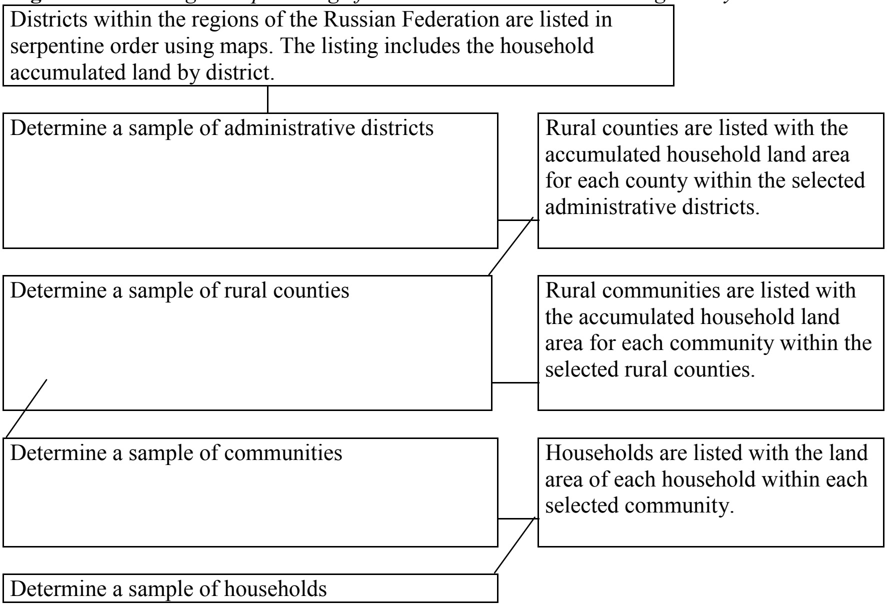
Stage II: A list of rural counties within each selected administrative district was compiled to include the number of households and accumulated land by county.

Stage III: A list of rural communities within each selected county was compiled to include the number of households and accumulated land by community.

Stage IV: List of households within each community complied to include hectares operated by each rural household.

At each of the four stages, samples were selected with probabilities proportional to size (the size measure being the accumulated land operated by rural households). At Stage I, 25% of administrative districts were selected, at Stage II, 15% of rural counties within selected districts were selected followed by 10% of rural communities chosen within each selected county at Stage III, and finally, at Stage IV the household sample size was determined based on the initially set level of 0.1% of the total household population of this region.

Figure 1: *Four Stage Sample Design for the Rural Household Farming Survey*



On the regional level, the data were summarized using the following estimators:

1) An average estimator (the population mean);
2) A probability estimator (probabilities of selection were utilized at each of the four stages);
3) A ratio estimator (ratio to land calculated at household level); and
4) A ratio estimator (the accumulated land of households selected within each district relative to the total household land of the district).

Once the data are summarized, the survey results are compared to the results of previous censuses, local administrative data, information from land resource and management committees as well to the official figures on crop yield, livestock productivity, fodder consumption per livestock and poultry head as reported by agricultural enterprises, and are checked for compliance with the current local regulations on plant and animal health.

The national private household farming totals are estimated based on the Rural Household Farming Survey results (for individuals living in the rural area) and ad-hoc data (specialized surveys conducted between censuses) on agricultural commodities produced at individual and collective fruit and kitchen gardens (for individuals living in the urban area.)

5.3 Private Farm Survey

Statistical data on agricultural activity of private farms are collected through regular sample surveys and periodical censuses. The main goal of these statistical efforts is to provide information on crop and livestock production, hectares planted to agricultural crops, livestock inventory, and poultry inventory.

A sample of private farms is developed using either single-purpose systematic or stratified sample selection procedures. In regions with small number of farm operations producing targeted types of agricultural commodities (less than 200 farms), systematic sampling proves to be the most efficient procedure. Stratified sampling may be successfully used when the universe is large enough (over 200 farms when the homogeneity of item values may be achieved due to a high number of observation units in each stratum.

The Private Farm sub-register serves as a sampling frame for the purpose of sample selection procedure. The universe includes farm operations producing the targeted agricultural commodities.

The universe is developed in the following way for crop and livestock surveys: for the crop growing category - a set of private farms with hectares planted to crops on June 1 is stratified by cropland area; for the livestock category – a set of private farms with cattle, hogs and pigs, sheep and goats, poultry, reindeers on January 1 is stratified by the relative measure of livestock inventory. The farm operation serves as a sample unit and an observation unit.

For the economic survey a list of farm operations within each region was compiled and stratified by values of the target item (planted area for the crop growing category, and livestock and poultry inventory for the livestock category), serves as a sample frame. The list was first grouped by the cropland area and then further sub-grouped by the relative measure of livestock inventory.

The results of the private farm sample were extended to the universe by strata based upon farm averages.

6. Changing Statistical System – Statistical Issues

In the Russian Federation's new economy, surveys are now used on a wider more sophisticated scale. Making this major transformation to a "statistical system" was not a trivial experience. While it is generally understood that sampling methodology should be embraced, introducing new, sophisticated, survey sampling procedures in a vast, diverse and changing country such as Russia generated new issues to be overcome. Some of the issues are outlined below.

- **Develop a completely new "survey system" to acquire agricultural statistical information for Russia's new economy under extremely tight time constraints**

  With the transition from a central planned economy to a market economy, GOSKOMSTAT of Russia had to act quickly to develop a series of sophisticated, statistical surveys that would specifically monitor the changing agricultural situation by type of operation – particularly for Private Households and Private Farms. Results from these surveys, along with reported agricultural information for enterprises, allowed GOSKOMSTAT of Russia to develop the necessary statistical reports on the changing agricultural situation.

- **Educate regional managers and staff on advantages and principles of statistical surveys**

  A number of regional staff members considered a census as the "only" reliable way to collect information and were reluctant to change to sample surveys. GOSKOMSTAT of Russia through training and seminars (New Statistical Procedures Seminar – 1996), information sharing, and pilot surveys had to convince the regional staff that statistical sample surveys were valid and cost effective.

- **Develop new statistical approaches to survey methodology and sample design**

  GOSKOMSTAT of Russia realized that more sophisticated statistical techniques and designs would be needed. One of the principle directions of collaboration with NASS was to raise the level of quality and sophistication of their survey procedures and sample designs. In the future, specific attention will be given to economic statistics with general methodology training and development of an economic survey for agricultural enterprises.

- **Devote resources (time and staff) into building registers**

  Building and developing a register of private farms to serve as a sampling frame for the purpose of selecting samples took considerable effort, thought and resources. Consideration was given to the targeted unit for sampling, the completeness of the register, the maintenance of the register (omission or duplication issues) and how quickly it was changing, information available on each entry that was needed for the sample design, and if the register was accessible and electronically maintained. The initial source of the register for private households was the existing Land Tax list maintained by the Land Valuation Committee. Regional statistical offices worked with the tax list for use as a sampling frame. Tax lists were often not electronically maintained.

- **Provide training to staff at Federal and Regional offices**

  GOSKOMSTAT of Russia collaborated with NASS in conducting a series of training seminars in Russia for staff members. Additionally, Russian delegates visited NASS headquarters in Washington DC and several state statistical field offices on several occasions for briefings on survey methodology and procedures. Statistical training will continue in both Russia and the United States. To enhance GOSKOMSTAT of Russia's ability to process survey data, plans include training in data analysis and the use of a statistical software package.

- **Selection and training of interviewers**

  Special attention was given to data enumeration techniques, particularly in establishing trust with the respondent to gain their cooperation in order to receive reliable information. Issues to consider were confidentiality in use of data, mandatory versus voluntary reporting, and potential impact of data results on the respondent. Russian farmers tendency to underreport production because of fear of increased taxes presented a challenge for interviewers.

- **Conduct pretests of survey process and procedures**

  It is critical when developing a new system to thoroughly pre-test the entire survey system. This includes testing the procedures, the questionnaire content and flow, software applications and hardware capabilities. GOSKOMSTAT of Russia conducted pretests in the Ryazan, Kirov, Penza, and Tver regions.

- **Dealing with limited technical resources**

  Regional GOSKOMSTAT offices faced serious shortages of computers and technical office supplies. To insure development of successful procedures, new desk-top computers were furnished to the pilot regional offices of Tver and Ryazan.

- **Efficient utilization of limited staff and financial resources**

  Cost efficiencies gained by sampling as compared to a census are particularly appealing to GOSKOMSTAT of Russia in a time when the need for information is increasing and staff and financial resources are declining.

7. Conclusion

The timely dissemination of survey results was an essential conclusion of the process for the surveys of rural households and private farming. For a population in transition, such as the Russian Federation, the value of agricultural and economic surveys for policy makers and others is to obtain accurate and timely statistical information. GOSKOMSTAT of Russia has produced the following publications in Russian and English: *Private Household Farming in Russia* (1999) and *Agricultural Activity of Private Farms in Russia* (2000).

GOSKOMSTAT of Russia and NASS will continue progress in improving the agricultural statistical system in Russia.

References

Private Household Farming in Russia, State Committee of the Russian Federation on Statistics, Moscow, Russia, 1999

Agricultural Activities of Private Farms in Russia, State Committee of the Russian Federation on Statistics, Moscow, Russia, 2000

Statistical Aspects of a Census

Carol C. House
National Agricultural Statistics Service, U.S. Department of Agriculture
3251 Old Lee Highway, Room 305 -- Fairfax, VA 22030-1504 USA
chouse@nass.usda.gov

Keywords: census, coverage, non-response, error, frames, imputation, disclosure

1. Introduction

In this paper the author will provide a basic overview of the statistical aspects of planning, conducting and publishing data from a census. The intent of the paper is to demonstrate that most (if not all) of the statistical issues that are important in conducting a survey are equally germane to conducting a census.
In order to establish the scope for this paper, we begin by reviewing some basic definitions. Webster's New Collegiate Dictionary defines a "census" to be "a count of the population and a property evaluation *in early Rome"*. Although particularly appropriate to quote at the *CAESAR* conference, we will want to utilize a broader definition. The International Statistical Institute (ISI) in its Dictionary of Statistical Terms defines a census to be "the complete enumeration of a population or group at a point in time with respect to well-defined characteristics". This definition is more useable. We now look at the term "statistics" to further focus the paper. Again from ISI we find that statistics is the "numerical data relating to an aggregate of individuals; the science of collecting, analyzing and interpreting such data." Together these definitions render a focus for this paper -- those issues germane to the science and/or methodology of collecting, analyzing and interpreting data through what is intended to be a complete enumeration of a population at a point in time with respect to well-defined characteristics. Further, because of the nature of the CAESAR conference, this paper will direct its discussion to agricultural censuses. Important issues include the (sampling) frame, sampling methodology, non-sampling error, processing, weighting, modeling, disclosure avoidance, and data dissemination. This paper touches on each of these issues as appropriate to the paper's focus on censuses of agriculture.

2. Frame

Whether conducting a sample survey or a census, a core component of methodology is the sampling frame. The frame usually consists of a listing of population units, but alternatively it might be a structure from which clusters of units can be delineated. For agricultural censuses, the frame is likely to be a business register or a farm register. Alternatively it might be a listing of villages from which individual farm units can be delineated during data collection. The use of an area frame is a third common alternative. Often more than a single frame is used for a census. Papers presented at the

Agricultural Statistics 2000 conference highlight the diversity of sampling frames used for agricultural censuses (Sward, et. al., 1998; Kiregyera, 1998; David, 1998).
There are three basic statistical concerns associated with sampling frames: coverage, classification and duplication. These concerns are equally relevant whether the frame will be used for a census or sampled for a survey.

2.1 Coverage

Coverage deals with how well the frame fully delineates all population units. The statistician's goal should be to maximize coverage of the frame and to provide measures of under-coverage. For agricultural censuses, coverage often differs by size of farming operation. Larger farms are covered more completely, and smaller farms less so. Complete coverage of smaller farms is highly problematic, and statistical organizations have used different strategies to deal with this coverage problem.
The Australian Bureau of Statistics (Sward, et. al., 1998) intentionally excludes smaller farms from their business register and census of agriculture. They focus instead on production agriculture, and maintain that their business register has good coverage for that target population. Statistics Canada (Lim, et. al., 2000) has dropped the use of an area frame as part of its census of agriculture, and is conducting research on using various sources of administrative data to improve coverage of its farm register. Kiregyera (1998) reports that a typical agriculture census in Africa will completely enumerate larger operations (identified on some listing), but does not attempt to enumerate completely the smaller operations because of the resources required to do so. Instead they select a sample from a frame of villages or land areas, and delineate small farms within the sampled areas for enumeration. In the United States, the farm register used for the 1997 Census of Agriculture covered 86.3% of all farms, but 96.4% of farms with gross value of sales over $10,000 and 99.5% of the total value of agricultural products. The U.S. uses a separate area sampling frame to measure under-coverage of its farm register, and has published global measures of coverage. They are investigating methodology to model under-coverage as part of the 2002 census and potentially publish more detailed measures of that coverage.

2.2 Classification

A second basic concern with a sampling frame is whether frame units are accurately classified. The primary classification is whether the unit is, in fact, a member of the target population, and thus should be represented on the frame. For example, in the U.S. there is an official definition of a farm: operations that sold $1,000 or more of agricultural products during the target year, *or would normally sell that much*. The first part of the definition is fairly straightforward, but the second causes considerable difficulty with classification.
Classification is further complicated when a population unit is linked with, or owned by, another business entity. This is an ongoing problem for all business registers. The statistician's goal is to employ reasonable, standardized classification algorithms that are consistent with potential uses of the census data. For example, a large farming operation may be a part of a larger, vertically integrated enterprise which may have holdings under semi-autonomous management in several dispersed geographic areas. Should each geographically dispersed establishment be considered a farm, or should the enterprise be considered a single farm and placed only once on the sampling frame? Another example is when large conglomerates contract with small, independent farmers

to raise livestock. The larger firm (contractor) places immature animals with the contractee who raises the animals. The contractor maintains ownership of the livestock, supplies feed and other input expenses, then removes and markets the mature animals. Which is the farm – the contractor, the contractee, or both?

2.3 Duplication

A third basic concern with a sampling frame is duplication. There needs to be a one-to-one correspondence between population units and frame units. Duplication occurs when a population unit is represented by more than one frame unit. Similar to misclassification, duplication is an ongoing concern with all business registers. Software is available to match a list against itself to search for potential duplication. This process may eliminate much of the duplication prior to data collection. Often it is important in a census or survey to add questions to the data collection instrument that will assist in a post-collection evaluation of duplication. In its 1997 Census of Agriculture, the U.S. conducted a separate "classification error study" in conjunction with the census. For this study, a sample of census respondents was re-contacted to examine potential misclassification and duplication, and to estimate levels of both.

3. Sampling

When one initially thinks of a census or complete enumeration, statistical sampling may not seem relevant. However, in the implementation of agricultural censuses throughout the world, a substantial amount of sampling has been employed. David (1998) presents a strong rationale for extensive use of sampling for agricultural censuses, citing specifically those conducted in Nepal and the Philippines. The reader is encouraged to review his paper for more details. This paper does not attempt an intensive discussion of different sampling techniques, but identifies some of the major areas where sampling has (or can be) employed.

Reducing costs is a major reason that statistical organizations have employed sampling in their census processes. We have already discussed how agricultural censuses in Africa, Nepal, and the Philippines have used sampling extensively for smaller farms. Sampling may also be used in quality control and assessment procedures. Examples include: conducting a sample survey of census non-respondents to assist in non-response adjustment; or conducting a specialized follow-up survey of census respondents to more carefully examine potential duplication and classification errors. The U.S. uses a sample survey based on an area frame to conduct a coverage evaluation of its farm register and census. It may be advantageous in a large collection of data to sub-divide the population and use somewhat different questionnaires or collection methodologies on each group. Here again is a role for sampling. For example, in order to reduce overall respondent burden some organizations prepare both aggregated and detailed versions of a census questionnaire and use statistical sampling to assign questionnaire versions to the frame units. Alternatively sampling may facilitate efforts to evaluate the effect of incentives, to use pre-census letters as response inducements, or to examine response rates by different modes of data collection.

4. Non-sampling Error

Collection of data generates sampling and non-sampling errors. We have already discussed situations in which sampling, and thus sampling error, may be relevant in census data collection. Non-sampling errors are always present, and generally can be expected to increase as the number of contacts and the complexity of questions increases. Since censuses generally have many contacts and fairly involved data collection instruments, one can expect them to generate a fairly high level of non-sampling error. In fact, David (1998) uses expected higher levels of non-sampling error in his rationale for avoiding complete enumeration in censuses of agriculture.

> "… [a census produces] higher non-sampling error which is not necessarily less than the total error in a sample enumeration. What is not said often enough is that, on account of their sizes, complete enumeration CA's [censuses of agriculture] use different, less expensive and less accurate data collection methods than those employed in the intercensal surveys."

Two categories of non-sampling error are response error and error due to non-response.

4.1 Response Error

The literature (Groves, 1989; Lyberg et. al., 1997) is fairly rich in discussions of various components of this type of error. Self-enumeration methods can be more susceptible to certain kinds of response errors, which could be mitigated, if interviewer collection were employed. Censuses, because of their large size, are often carried out through self-enumeration procedures. The Office of National Statistics in Britain (Eldridge, et al; 2000) has begun to employ cognitive interviewing techniques for establishment surveys much the same as they have traditionally employed for household surveys. They conclude that the "… use of focus groups and in-depth interviews to explore the meaning of terms and to gain insight into the backgrounds and perspectives of potential respondents can be very valuable …" They further conclude regarding self-administered collection that "…layout, graphics, instructions, definitions, routing etc. need testing." Kiregyera (1998) additionally focuses readers' attention on particular difficulties that are encountered when collecting information from farmers in developing countries. These include the "failure of holders to provide accurate estimates of crop area and production … attributed to many causes including lack of knowledge about the size of fields and standard measurement units, or unwillingness to report correctly for a number of reasons (e.g. taboos, fear of taxation, etc.)."
The statistician's role is fourfold: to understand the "total error" profile of the census, to develop data collection instruments and procedures that minimize total error, to identify and correct errors during post collection processing, and to provide, to the extent reasonable, measures of the important components of error.

4.2 Non-Response

The statistician's role in addressing non-response is very similar to his/her role in addressing response error: to understand the reasons for non-response, to develop data collection procedures that will maximize response, to provide measures of non-response error, and to impute or otherwise adjust for those errors.

Organizations employ a variety of strategies to maximize response. These include publicity, pre-collection contacts, and incentives. Some switch data collection modes between waves of collection to achieve higher response rates. Others are developing procedures that allow them to target non-response follow-up to those establishments which are most likely to significantly impact the estimates. (McKenzie, 2000)

A simple method for adjusting for unit non-response in sample surveys, is to modify the sampling weights so that respondent weights are increased to account for non-respondents. The assumption in this process is that the respondents and non-respondents have similar characteristics. Most often, the re-weighting is done within strata to strengthen the basis for this assumption. A parallel process can be used for censuses. Weight groups can be developed so that population units within groups are expected to be similar in relationship to important data items. All respondents in a weight group may be given a positive weight, or donor respondents may be identified to receive a positive weight. Weight adjustment for item non-response, although possible, quickly becomes complex as it creates a different weight for each item.

Imputation is widely used to address missing data, particularly that due to item non-response. Entire record imputation is also an appropriate method of addressing unit non-response. Manual imputation of missing data is a fairly widespread practice in data collection activities. Many survey organizations have been moving toward more automated imputation methods because of concerns about consistency and costs associated with manual imputation, and to improve the ability to measure the impact of imputation. Automating processes like imputation are particularly important for censuses because of the volume of records that must be processed.

Yost et. al. (2000) identify five categories of automated imputations: i) deterministic imputation – where only one correct value exists (such as the missing sum at the bottom of a column of numbers; ii) model-based imputation – use of averages, medians, ratios, regression estimates, etc. to impute a value; iii) deck imputation – a donor questionnaire is used to supply the missing value; iv) mixed imputation – more than one method used; and v) the use of expert systems. Many systems make imputations based on a specified hierarchy of methods. Each item on the questionnaire is resolved according to its own hierarchy of approaches, the next being automatically tried when the previous method has failed. A nearest neighbor approach based on spatial "nearness" may make more sense for a census, where there is a greater density of responses, than it would in a more sparsely distributed sample survey.

5. Post Collection Processing

Post collection processing involves a variety of different activities, several of which (imputation, weighting, etc.) are discussed in other sections of this paper. Here we will briefly address editing and analysis of data. Because of the volume of information associated with a census data collection, it becomes very important to automate as many of these edit and analyses processes as possible. Atkinson and House (2001) address this issue and provide several guiding principles that the National Agricultural Statistical Service is using in building an edit and analysis system for use on the 2002 Census of Agriculture: a) automate as much as possible, minimizing required manual intervention; b) adopt a "less is more" philosophy to editing, creating a leaner edit that focuses on critical data problems; and c) identify problems as early as possible.

Editing and analysis must include the ability to examine individual records for consistency and completeness. This is often referred to as “micro” editing or “input” editing. Consistent with the guiding principles discussed above, the Australian Bureau of Statistics has implemented the use of significance criteria in input editing of agricultural data. (Farwell and Raine, 2000) They contend that “… obtaining a corrected value through clerical action is expensive (particularly if respondent re-contact is involved) and the effort is wasted if the resulting actions have only a minor effect on estimates." They have developed a theoretical framework for this approach.
Editing and analysis must also include the ability to perform macro-level analysis or output editing. These processes examine trends for important subpopulations, compare geographical regions, look at data distributions and search for outliers. Desjardins and Winkler (2000) discuss the importance of using graphical techniques to explore data and conduct outlier and inlier analysis. Atkinson and House concur with these conclusions and further discuss the importance of having the macro-analysis tool integrated effectively with tools for user-defined ad-hoc queries.

6. Weighting

When one initially thinks of a census, one thinks of tallying up numbers from a complete enumeration, and publishing that information in a variety of cross tabulations that add to the total. This paper has already discussed a variety of situations in which weighting may be a part of a census process. In this section we focus on the interaction between weighting and the rounding of data values.
Many of the important data items collected in an agricultural census are intrinsically “integral” numbers, making sense only in whole increments (i.e. the number of farms, number of farmers, number of hogs, etc.). For these data, desirable characteristics of the census tabulation is to have integer values at all published levels of disaggregation, and to have those cells sum appropriately to aggregated totals.
The existence of non-integer weights creates non-integer weighted data items. Rounding each of the multiple cell totals creates the situation that they may not add to rounded aggregate totals. This issue can be addressed in one of several ways. In the U.S., the census of agriculture has traditionally employed the technique of rounding weights to integers, and then using these integerized weights. An alternative would be to retain the non-integer weights and round the weighted data to integers. A recent evaluation of census data in the U.S. (Scholetsky, 2000) showed that totals produced using the rounded weighted data values were more precise than the total produced using the integerized weights except for the demographic characteristics, number of farms, and ratio per farm estimates. A drawback to using rounded weighted data values is the complexity these procedures add to storing and processing information.

7. Modeling

Modeling can be effective within a census process by improving estimates of small geographic areas and rare subpopulations. Small area statistics is perhaps one of the most important products from a census. However, a number of factors may impact the census’ ability to produce high quality statistics at fairly disaggregate levels. The highly skewed distribution of data, which is intrinsic to the structure of modern farming, creates estimation difficulties. For example, many larger operations have production units which cross the political or geographic boundaries used in publication. If data are collected for the large operation and published as if the "whole" farm is contained within a single geographic area, this result will be an over-estimate of agricultural production within that area and a corresponding under-estimate within surrounding areas. Mathematical models may be used effectively to prorate the operation totals to appropriate geographic areas.

Census processes for measuring and adjusting non-response, misclassification, and coverage may produce acceptable aggregate estimates while being inadequate for use at the more disaggregate publication levels. Statistical modeling and smoothing methodology may be used to smooth the measures so that they produce more reasonable disaggregate measures. For example, for the 1997 Census of Agriculture the U.S. provided measures of frame coverage at the state level for farm counts for major subpopulations. They are evaluating several smoothing techniques that, if successful, may allow the 2002 census release to include coverage estimates at the county level instead of just state level, and for production data as well as farm counts.

Although a census may be designed to collect all information from all population units, there are many cases in which circumstances and efficiencies require that census data not stand alone. We have already discussed methodologies in which a separate survey may be used to adjust census numbers for non-response, misclassificaion and/or coverage. Sometimes sources of administrative data are mixed with census data to reduce respondent burden or data collection costs. Most often the administrative data must be modeled to make it more applicable to the census data elements. Alternatively, some census collection procedures utilize a "long" and "short" version of the questionnaire so that all respondents are not asked every question. To combine the data from these questionnaire versions may also require some type of modeling.

8. Disclosure Avoidance

The use of disclosure avoidance methodology is critically important in preparing census and survey data for publication. Disclosure avoidance can be very complex for agricultural census publications because of the scope, complexity and size of these undertakings. Disclosure avoidance is made more difficult by the highly skewed nature of the farm population. Data from large, or highly specialized, farming operations are hard to disguise, especially when publishing totals disaggregated to small geographic areas.

Disclosure avoidance is typically accomplished through the suppression of data cells at publication. A primary suppression occurs when a cell in a publication table requires suppressing because the data for the cell violates some rule or rules defined by the statistical agency. Typical rules include:

a) *threshold rule*: the total number of respondents is less than some specified number, i.e. the cell may be suppressed if it had fewer than 20 positive responses.
b) *(n,k) rule*: a small number of respondents constitute a large percentage of the cell's value, for example a (2,60) rule would say to suppress if 2 or fewer responses made up 60 percent or more of the cell's value.

c) *p-percent rule*: if a reported value for any respondent can be estimated within some specified percentage.

Secondary suppression occurs when a cell becomes a disclosure risk from actions taken during the primary suppression routines. These additional cells must be chosen in a way that provide adequate protection to the primary cell and at the same time make the value of the cell mathematically underivable.

Zayatz et al. (2000) have discussed alternatives to cell suppression. They propose a methodology that adds "noise" to record level data. The approach does not attempt to add noise to each publication cell, but uses a random assignment of multipliers to control the effect of the noise on different types of cells. This results in the noise having the greatest impact on sensitive cells, with little impact on cells that do not require suppression.

9. Dissemination

Data products from a census are typically extensive volumes of interconnected tables. The Internet, CD-rom, and other technical tools now provide statistical agencies with exciting options for dissemination of dense pools of information. This paper will discuss several opportunities to provide high quality data products.

The first component of a quality dissemination system is metadata, or data about the data. Dippo (2000) expounds on the importance of providing metadata to users of statistical products and on the components of quality metadata.

> "Powerful tools like databases and the Internet have vastly increased communication and sharing of data among rapidly growing circles of users of many different categories. This development has highlighted the importance of metadata, since easily available data without appropriate metadata could sometimes be more harmful than beneficial."

> "Metadata descriptions go beyond the pure form and contents of data. Metadata are also used to describe administrative facts about data, like who created them, and when. Such metadata may facilitate efficient searching and locating of data. Other types of metadata describe the processes behind the data, how the data were collected and processed, before they were communicated or stored in a database. An operational description of the data collection process behind the data (including e.g. questions asked to respondents) is often more useful than an abstract definition of the "ideal" concept behind the data."

The Internet has become a focal point for the spread of information. Web users expect: to have sufficient guidance on use; to be able to find information quickly, even if they do not know precisely what they are looking for; to understand the database organization and naming conventions; and to be able to easily retrieve information once it is found. This implies the need, at a minimum, for high quality web design, searchable databases, and easy to use print and download mechanisms. The next step is to provide tools such as interactive graphical analysis with drill-down capabilities and fully functional interactive query systems. Graphs, charts and tables would be linked, and users could switch between these different representations of information. Finally, there would be links between the census information and databases and websites containing information on agriculture, rural development, and economics.

10. Summary

Conducting a census involves a number of highly complex statistical processes. One must begin with a quality sampling frame, in which errors due to under-coverage, mis-classification and duplication are minimized. There may be opportunities in which statistical sampling will help bring efficiency to the data collection or facilitate quality control measurements. Non-sampling errors will be present, and the design must deal effectively with both response and non-response errors. Post collection processing should allow both micro and macro analysis. Census processing will probably involve weighting and some type of modeling. The dissemination processes should prevent disclosure of respondent data while providing useful access by data users.

References

Atkinson, D., House, C. (2001) A Generalized Edit and Analysis System for Agricultural Data, *Proceedings of the Conference on Agricultural and Environmental Statistical Application in Rome,* International Statistical Institute. The Netherlands.

David, I. (1998) Sampling Strategy for Agricultural Censuses and Surveys in Developing Countries, *Proceedings of Agricultural Statistics 2000, 83-95.* International Statistical Institute. The Netherlands.

DesJardins, D., Winkler, W. (2000) Design of Inlier and Outlier Edits for Business Surveys, *Proceedings of the 2nd International Conference on Establishment Surveys*, 547-556, American Statistical Association, Washington.

Dippo, C. (2000) The Role of Metadata in Statistics, *Proceedings of the 2nd International Conference on Establishment Surveys*, 909-918, American Statistical Association, Washington.

Eldridge, J., Martin, J., White, A. (2000) The Use of Cognitive Methods to Improve Establishment Surveys in Britain, *Proceedings of the 2nd International Conference on Establishment Surveys*, 307-316, American Statistical Association, Washington.

Farwell, K., Raine, M. (2000) Some Current Approaches to Editing in the ABS, *Proceedings of the 2nd International Conference on Establishment Surveys* 529-538, American Statistical Association, Washington.

Groves, R. (1989) *Survey Errors and Survey Costs*, John Wiley & Sons, New York.

International Statistical Institute (1990) *A Dictionary of Statistical Terms*, Published for the International Statistical Institute by Longman Scientific & Technical. Essex CM20 2JE England.

Kiregyera, B. (1998) Experiences with Census of Agriculture in Africa*, Proceedings of Agricultural Statistics 2000, 71-82,* International Statistical Institute. The Netherlands.

Lim, A., Miller, M., Morabito, J. (2000) Research Into Improving Frame Coverage for Agricultural Surveys at Statistics Canada, *Proceedings of the 2nd International*

Conference on Establishment Surveys, 131-136, American Statistical Association, Washington.

Lyberg, L., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., Trewin. D., Eds. (1997) *Survey Measurement and Process Quality*, John Wiley & Sons, Inc. New York.

McKenzie, R. (2000) A Framework For Priority Contact of Non Respondents, *Proceedings of the 2nd International Conference on Establishment Surveys*, 473-482, American Statistical Association, Washington.

Scholetzky, W. (2000*) Evaluation of Integer weighting for the 1997 Census of Agriculture*, *RD Research Report Number RD-00-01* National Agricultural Statistics Service. U. S. Department of Agriculture. Washington.

Sward, G., Hefferman, G., and Mackay, A. (1998) Experience with Annual Censuses of Agriculture, *Proceedings of Agricultural Statistics 2000*, 59-70, International Statistical Institute. The Netherlands.

Webster's New Collegiate Dictionary (1977) G. & C. Merriam Company, Springfield, MA.

Yost, M., Atkinson, D., Miller, J., Parsons, J., Pense, R., Swaim, N. (2000) Developing *A state of the Art Editing, Imputation and Analysis System for the 2002 Agricultural Census and Beyond--An unpublished staff report*, National Agricultural Statistics Service. U. S. Department of Agriculture. Washington.

Zayatz, L., Evans, T., Slanta, J. Using Noise for Disclosure Limitation of Establishment Tabular Data, *Proceedings of the 2nd International Conference on Establishment Surveys*, 877-886, American Statistical Association, Washington.

6 June 2001

Invited Paper Session

TIMELINESS/ACCURACY/COMPLETENESS OF AGRICULTURAL STATISTICS

Chair: F. A. Vogel

Using Administrative Data for Census Coverage

Claude Julien
Statistics Canada, Ottawa, ON, Canada, K1A 0T6
e-mail: claude.julien@statcan.ca

Abstract: Statistics Canada will be conducting a concurrent Census of Population and Census of Agriculture in May 2001. Census enumerators will again visit every household in Canada with drop-off questionnaires. This collection methodology has been effective in terms of coverage of household-based agriculture operations. In recent censuses, special collection procedures have been put in place to cover very large farms, multi-operation farms, specific farm types and certain areas in the country. The last census conducted in 1996 gave indications that these procedures may not be able to in fully keep up with the rapid structural changes in the agriculture sector. This paper describes research conducted into using administrative data to maintain the census coverage at its historically high level and to assist in evaluating the coverage.

Key words: Coverage, Census, Administrative data

1. Background

In Canada, concurrent national demographic and agricultural censuses have been conducted every five years since 1956. The next censuses will be conducted on May 15, 2001. As in previous censuses, agricultural operators will be required to complete both the Census of Agriculture and Census of Population questionnaires.

The Census of Agriculture produces a snapshot of Canadian agriculture by providing statistics at national, provincial and subprovincial levels on crop areas, numbers of livestock, numbers and value of farm machinery, farm operating expenses and receipts, purchases of capital assets, weeks of paid labour and land management practices (Statistics Canada, 1997). Conducting the Census of Agriculture jointly with the Census of Population helps streamline collection procedures and saves millions of dollars. This method is effective in terms of farm coverage as long as agriculture remains a household-based industry. This was still the case according to the 1996 Census when 88% of the 276,000 farms enumerated were classified as a sole-proprietorship or a partnership and 86% of the 385,000 farm operators reported residing on their farm.

In May 2001, Census representatives will deliver Census of Population questionnaires to every household in their respective enumeration area. A Census of Agriculture questionnaire will also be dropped off if it is reported or there is evidence that someone in the household operates an agricultural operation. Census of Agriculture respondents are asked to mail back their completed questionnaires to field collection offices where they will be checked and corrected for completeness and accuracy. Afterwards, the questionnaires will be sent to the central processing office where they are captured, processed, validated, certified and tabulated for public data release. This series of operations will take one year from Census Day to data release.

Data quality evaluation studies were performed on every Census of Agriculture between 1956 and 1981 (Statistics Canada 1984). In 1956, 1961 and 1966, quality check surveys were conducted immediately following the census. Areas sampled from across the country were re-enumerated by census representatives in an attempt to identify coverage and response errors. In 1971, a post-censal agriculture survey was carried out and data from this survey were used to evaluate the quality of the census counts. In 1976 and 1981, annual agricultural probability surveys, that were conducted one month after the census, provided aggregate level estimates for comparison to census totals. The samples for the surveys were drawn from an area frame supplemented by a list frame of farms that were major producers of certain important commodities.

Since 1986, the list frame component of the agriculture surveys steadily increased to the point of dropping the area frame component altogether after the 1996 Census. Consequently, it became too expensive to conduct a coverage evaluation study using a proper area frame sample. According to the last study conducted in 1981, the Census of Agriculture had enumerated 96.8% ± 2.6% of all farms. Section 2 of this paper describes changes in the agriculture industry that have made it relevant to re-assess this situation. Section 3 introduces Statistics Canada's Farm Register. Section 4 describes the methodology and the results of a study that evaluated the feasibility of using tax data to improve the coverage of the Register. Section 5 presents how the Register will be used to maintain and evaluate census coverage. Section 6 concludes with a list of projects, under consideration, related to coverage improvement.

2. Census coverage

Since 1981, the focus has been on maintaining a high level of coverage. Several supplemental operations have been added in recent censuses to complement the drop-off / mail-back collection operation.

A question asking if anyone in the household operates an agricultural operation is asked on the Census of Population questionnaire to identify and follow-up on farm operators who were not given an agriculture questionnaire when the drop-off took place.

Community pastures, research stations, multi-operation corporations, large farm corporations and unique one-of-a-kind operations are identified prior to census enumeration. They are enumerated via specially negotiated collection arrangements directly from the central processing office. In areas where agriculture activity is scarce, a list of agricultural operations enumerated in the previous census is provided to the census representative to assist them in their enumeration operation.

Once the census questionnaires are returned to the central processing office, they are matched against a list of the larger farms enumerated in the previous census. The farms on the list that do not match to a census questionnaire are followed-up by telephone from the central office. Respondents are asked if they still operate a farm and to complete a census questionnaire when warranted. This operation must be carried out very quickly and only on a limited number of large farms in order to include their census questionnaire within the tight data processing and tabulation schedule. Prior to the 1996 census, this post-enumeration operation contributed between 500 and 1,000 additional farms (i.e. approximately 0.2% of all census farms).

2.1. Recent challenges in maintaining census coverage

Several changes to the face of agriculture that have occurred or are occurring are posing a definite challenge in maintaining census coverage at its historical high levels. An increase in the turnover in farm ownership, a higher incidence of land rental, an increase in the number of non-household based farm operations (large, non-family corporate farms), farm operators who do not reside on their operations and a non-traditional look to farming have made it more difficult to correctly identify all farms. A decrease in the number of farms in certain areas, as well as a decrease in the proportion of farm operators in rural areas may also have an impact on the level of census coverage. Finally, poorer economic conditions in some sectors of agriculture may mean that more respondents are working off the farm thus reducing the possibility of making a personal contact at the questionnaire drop-off operation.

The impact of these changes came to light in the 1996 Census when over 5,000 farms were identified and added during the post-enumeration operation. These farms represented 2% of all farms enumerated in Canada, 1.7% of field crop acreage, 2% of cattle inventory and 5% of hog inventory. The changes listed above appeared to have contributed to the increase in the number of farms added in that only 70% of them were managed by a resident operator compared to 86% of the farms enumerated in the drop-off/mail-back operation. The results of the 1996 Census gave a signal that (1) the use of an independent list to supplement the census enumeration had to be expanded and (2) census coverage had to be re-assessed.

3. Farm register

Due to changes in the agriculture population, it is increasingly becoming a challenge to fully enumerate all farms (demographic census) and all farming activity (economic census) using the traditional household based collection procedures in the Census of Population. However, the time and costs savings of conducting the two censuses together still greatly out-weigh the shortcomings in terms of coverage. In order to deal with these shortcomings, the 2001 Census of Agriculture must rely, more than in previous censuses, on the contribution of a good independent list of farms to maintain excellent coverage. This list will be extracted from Statistics Canada's Farm Register.

The Farm Register is a central database on which administrative data on Canadian agricultural operations and operators is stored. The Register is an essential tool for all agricultural surveys at Statistics Canada. Approximately 15 agricultural surveys, comprised of over 50 survey occasions are conducted every year. All the surveys use list frames extracted from the Register. The Register goes through a complete update after each Census of Agriculture. Between censuses, farm records are updated by survey feedback that provides changes to a farm's operating status (in or out-of-business) or a farm's operators (change-in-operators). All agricultural surveys at Statistics Canada are farm based (not land based) and are conducted by telephone, using Computer Assisted Telephone Interviewing technology. This framework has limitations in recording all changes to the farms contacted. As a result, for some farms, it is not known whether they are in operation or who is operating them (status unknown).

The Farm Register went through a major review following the 1996 Census when information collected on 276,362 farms was used to update the farm records. Since then, only 2,440 records, identified from various sources, have been added to the Register. Survey feedback is the primary source of updates to the Register. Since the major review, over 159,000 farm records (57% of the Register) have been selected for at least once survey and have been subject to survey feedback. Over 32,000 (20%) of the selected farms have reported a sale, rental or out-of-business change (the detailed distribution is provided in Table 1). A source of information that can identify farm activity in all commodities and all provinces is required to offset the attrition due to these changes.

Table 1: *Statistics on updates to the Farm Register*

Number of farms at the 1996 Census		276,362
Number of farm records added since the census		2,440
Number of farms selected for a survey since the census		159,510
Number of changes in farm status		32,604
	Out-of-business	4,398
	Status unknown	8,836
	Change-in-operators	19,370

From the early 1970s to the mid-1990s, Statistics Canada had conducted an annual survey based on an area frame to collect agricultural data. The area frame sample was combined with list frame samples to produces estimates covering the whole industry. At the beginning, the area frame accounted for most of the coverage. As the development of the list frame progressed, the purpose of the area frame was reduced to locate farms that were new since the last Census, or missed by the Census. It eventually became very expensive on a "per-farm" basis and was dropped following the 1996 Census. This has left the Register with no major source of updates besides on-going surveys and the Census of Agriculture.

4. Farm update survey

In order to offset the attrition on the Farm Register and the loss of the area frame, Statistics Canada conducted a Farm Update Survey to assess the feasibility of using income tax data to identify new farm operators and new farm operations between censuses as well as to update existing farms on the Farm Register (Lim, Miller, Morabito, 2001).

4.1. Tax data

In terms of a statistical universe, Canadian income tax data is divided in two major components, individuals (T1 universe) and corporations (T2 universe). The research focussed on the T1 universe because it covers over 75% of all farming activity in terms of revenue and the T2 universe is in transition towards a fully electronic reporting system and was not yet suited for the type of research conducted. On the plus side, tax data are available annually, they cover most of the agricultural activity in all commodities and all geographic areas and they come in a standard electronic format that is easy to use. On the negative side, tax data are composed of units (individuals) that are

different than the reporting unit used in the agriculture census and surveys (farms). Furthermore, tax data are available one year after the reference year. For example tax data for reference year 1999 are available only in December 2000.
Every year, approximately 450,000 individuals report some revenue from the sales of agricultural products in Canada on their income tax forms. This is far more than the approximately 250,000 individual or family farms (not incorporated) enumerated in the 1996 Census, and operated by 340,000 to 360,000 individuals. The objectives of the Farm Update Survey were: (1) to reduce the tax universe by transforming it from a population of individuals to that of likely partners of potential farm operations and (2) to identify the actual farm operations and add them to the Farm Register.
The first objective listed above was investigated by grouping individual tax records together into "fiscal farms", as defined below, and assessing their likelihood of actually being a farm. The second objective was investigated by matching the fiscal farms to the Farm Register and contacting the ones that did not match.

4.2. Methodology

The Farm Update Survey was comprised of four steps: forming groups of individual tax records, assigning a score to each group according to its size and its likelihood of actually operating a farm, looking up these groups on the Register and conducting a survey on the groups that were not found on the Register.

Step 1 GROUPING The first step of the Farm Update Survey (FUS) was to group tax records of individuals into likely partners of a potential farming operation. This was carried out by using various combinations of the family name, the address and the reported farm income of the individuals, as presented in Table 2. For example, individuals living in the same large area (Province) and reporting the exactly same very large farm income (over $250,000) were grouped into the same fiscal farm. A study on the robustness of the grouping rules indicated that over 97% of the individuals grouped into a fiscal farm in one reference year are grouped into the same fiscal farm in the subsequent reference year.

Table 2: - *Grouping rules*

Province	Postal code	Address	Last name	Reported Farm Income
✔				✔(>$250,000)
✔		✔		✔(>$0)
✔				✔(>$10,000)
✔	✔		✔	✔(>$0)
✔		✔	✔	

Step 2 SCORING The next step was to assess the likelihood that a fiscal farm was in fact a farm. To this end, we derived a score for each fiscal farm by combining the following variables: the average reported farm income among the individuals in the fiscal farm, the postal code of the individuals (rural or urban) and an indicator of the presence of one of the individuals registered to a farm income program. This score function enables the prioritization of the fiscal farms, with the expectation that the fiscal farms with the lower scores were less likely to be farms.

Step 3 MATCHING The purpose of the study was to identify farm operators and farm operations that were not already on the Farm Register. Once the fiscal farms were created and assigned a score their associated individuals were matched to the operators present on the Register using their name, date of birth and address. A fiscal farm was considered as present on the Register (and, thus, an existing farm) when at least one of its individuals was found on the Register. The fiscal farms that did not match to the Register made up the survey population for the last step.

Step 4 DATA COLLECTION AND PROCESSING The next step was to select the fiscal farms that were the most likely to be actual farms of significant size and to contact them. The purpose of the contact was two-fold: to determine if the individuals involved in the fiscal farms were actually farming and to get more up-to-date information on the operators of the farm(s) in question. The latter information was used to look up the fiscal farm once more on the Register to correct errors in the initial search due to erroneous or incomplete information on the Register or tax data, as well as to improve upon the limitations of the matching processes involved.

The operations described above produce two main results. First, they identify farm operators and farm operations that are not present on the Register. These operations are then added to the Register to improve its coverage and to reduce its attrition. Second, fiscal information on a significant proportion of the farm records is available and can be used to update their status on the Register.

4.3. Results

The methods and operations described above were carried out using tax data for reference year 1997 and Farm Register information extracted in early 1999. The study started with a tax universe of 455,250 individuals who reported some farm income. The individual tax records were grouped into 327,028 fiscal farms. The fiscal farms were assigned a score from 0 to 468. Their distribution is presented in Table 3. Matching the tax individuals to register operators found 243,458 fiscal farms present on the Register, leaving 83,570 fiscal farms to investigate. Table 3 indicates that fiscal farms with a score between 50 and 199 were more likely to be present on the Register (and, thus, actually farms). It was decided to investigate the fiscal farms with a score of 50 or more. This resulted in a sample of 11,029 fiscal farms that were contacted and processed. The completion rate was 78%; these farms were coded as either non-farm, farm already on the Register or new farm. Overall, 41% of the fiscal farms were reported as non-farms and 38% as new farms. It is particularly noteworthy to point out that the non-farm rate decreases as the score increases, thus indicating how well the score function succeeds at identifying farm operations.

Table 3: *Results from the Farm Update Survey*

Grouping and scoring		Matching		Collection and processing		
Score	Fiscal farms (counts)	Present on register	Not on register (count)	Non-farm	Farm on register	Farm added to register
000-024	146,280	61%	57,247	N/A	N/A	N/A
025-049	52,419	72%	14,710	N/A	N/A	N/A
050-099	58,380	90%	5,996	49%	14%	37%
100-149	56,575	93%	3,879	39%	21%	40%
150-199	11,709	88%	1,378	25%	37%	38%
200-249	1,255	80%	254	21%	44%	35%
250+	410	74%	106	18%	49%	33%
All	327,028	74%	83,570	41%	21%	38%

The study identified 3,280 farms to be added to the Register based on a proportion of one tax year. That is almost twice as many as what was added to the Register since the last census from any other source. More farms could be added to the Register by investigating the fiscal farms with a score of less than 50, but it is expected that more than 50% of them will be non-farms and, as the results discussed below indicate, their contribution to improving the coverage of agriculture activity is likely to be insignificant.

4.4. Coverage of farming activity by the farm register

The analysis of fiscal farms according to their reported farm income provides an indication of how much farming activity is covered by the Farm Register before and after the addition of new farms. After the grouping and matching operations described above, 71% of the fiscal farms were found on the Register and they accounted for 83% of the total reported farm income. Table 4 illustrates how the match rate was better among fiscal farms in the $50,000 - $3,000,000 compared to the other size categories.
The collection and data processing steps improve the match rate by: (1) identifying non-farming activity (leading to a reduction of the fiscal farm population), (2) matching more fiscal farms to the Register and (3) adding them to the Register as new farms. As a result, the percentage of fiscal farms present on the Register increased to 77%. These farms account for 97% of the total reported farm income.

Table 4: *Presence of fiscal farms on the register*

Reported farm income	Before the FUS	After the FUS
$1-$30K	62%	62%
$30K-$50K	87%	96%
$50K-$250K	94%	99%
$250K-$1M	95%	99%
$1M-$3M	93%	99%
$3M+	71%	99%
All fiscal farms	71%	77%
Total reported income	83%	97%

4.5. Preparing the farm register for census coverage

The methods developed and implemented for the Farm Update Survey were used to further update the Farm Register before it is used to assist in achieving excellent coverage in the 2001 Census. Tax data for reference year 1999 provided 436,123 individuals who reported some farm income on their tax returns. The individuals who were members of FUS fiscal farms that were not involved in farming activity (non-farms) in the FUS were dropped from the population. The rest of the population was grouped into 303,319 fiscal farms. The FUS methodology was further improved by obtaining an industrial classification code for over 50% of the fiscal farms from Statistics Canada's Business Register. This code was incorporated into the scoring function to derive a better *likelihood indicator*, as presented in Table 5. Based on data from the FUS, this likelihood indicator should achieve a better success at identifying actual farms among the fiscal farms than the score used in the FUS. The matching operation was carried out and it identified 71,939 fiscal farms that are not on the Register.

Table 5: *Fiscal farms from 1999 tax data that are not on the register*

Likelihood indicator	Estimated Likelihood*	Number of fiscal farms			
		>$30K	$10K-$30K	<$10K	Total
0	29%	905	2,876	15,118	18,899
1	54%	3,044	8,360	33,219	44,623
2	64%	1,686	2,209	4,468	8,363
3	87%	54	-	-	54
Total		5,689	13,445	52,805	71,939

* Estimated percentage of actual farms based on FUS results obtained for fiscal farms that reported $30,000 or more. This estimate cannot be applied to fiscal farms reporting less than $30,000.

In an attempt to strike a balance between time, resources and achieving good economic and demographic coverage, it was decided to add the fiscal farms with a likelihood of at least 1 and that reported a farm income of $30,000 or more in 1999 (4,784 fiscal farms). According to the results achieved from the Farm Update Survey, this should allow the Farm Register to cover at least 97% of the farming activity conducted by the individuals who reported some farm income in 1999. Depending on the initial indications on coverage observed during census processing and the resources available at that time, a sample of fiscal farms that reported less than $30,000 could be selected to improve the assessment of the coverage of the farm population.

Besides adding farms to the Register, the fiscal farms can be used to update the status of existing records on the Register. In fact, there are farm records coded as out-of-business prior to 1999 and that match to a 1999 fiscal farm. That is, the farm is coded as out-of-business since 1998 or before while its operator(s) reported farm income in 1999. This apparent contradiction could be accurate, as it could be that the operators resumed farming activity or the farm record may have been coded erroneously. These records will be reactivated prior to using the Register for census coverage purposes. Furthermore, farm records with a status unknown since 1998 or before and that match to a 1999 fiscal farm provides an indication that the farm is likely still in-business.

5. Coverage in the 2001 census of agriculture

An independent list of farms will be extracted from the Farm Register and used to assist the Census of Agriculture team in achieving a high level of coverage. This list will also be used to assess the census coverage.

5.1. Coverage follow-up operation

A snapshot of the Farm Register will be extracted on Census Day, May 15, 2001. This snapshot will include all farm records except the special agriculture operations that will be enumerated through specially negotiated arrangements and farm records that are coded as out-of-business or change-in-operators. The larger farms, depending on the commodity and the province, identified by the previous census or more current survey information will be flagged as "must-get" farms.

Once the census questionnaires are received and captured in the central processing office, they will be matched to the farm records on the snapshot. At a certain point in data processing, the must-get farms from the snapshot that will not have been matched to a census record (e.g. may not have completed a census questionnaire) will be contacted by telephone to complete a census questionnaire (to reduce undercoverage). The more up-to-date information on the contacted must-get farm will be matched to the census records to determine whether it should be allowed into the census data processing stream (to guard against duplication). This operation will be carried out in a very short period of time, between one and three weeks depending on the province. This time constraint restricts this operation to only the larger farms in a given province or specific commodity. An attempt will be made to contact as many farms as possible in the time available and, thus, to focus on achieving the best coverage of the farming economy. This is important not only to the census, but also to the next intercensal survey program that will be designed based on the census data.

5.2. Coverage assessment

The regular census collection operation and the ensuing coverage follow-up operation will not account for all must-get farms, mainly due to methodological, operational and time limitations. The remaining must-get farms along with the other snapshot records that will not have been flagged as must-get farms in the first place will make up the survey population for the coverage evaluation study. This study will select a sample of farms that do not match to a census record. The sample design will cover all farms in the survey population, big and small. The selected farms will be contacted much like fiscal farms were contacted in the Farm Update Survey. Information will be collected to determine whether the farm was in-operation on Census Day and to obtain more information on the farm and its operators to accurately determine whether the farm was enumerated or not. The data collection operation will take place after that of the coverage follow-up operation. Very little commodity information will be collected, as the record will not be included in the regular census data processing. The coverage evaluation study will thus focus on assessing the coverage of the farm population (or demography) only.

6. Conclusion

The Canadian Census of Agriculture has historically achieved excellent coverage. In recent censuses, the Census of Agriculture team has been pro-active in implementing several operations to maintain this coverage in light of changes in the agriculture industry. For the 2001 Census, an extensive operation will be carried out to prevent potential undercoverage in a timely manner. This operation will rely on a list of farms extracted from Statistic Canada's central Farm Register. This list will also be used to assess the coverage actually achieved after the regular data collection and the timely follow-up operations are completed.

Research on using tax data to improve the coverage of the Register has added 8,000 farms (3% of the current Register) and updated the status of other farm records that otherwise could have been considered as out-of-business. These updates have increased the Register coverage to over 97% of the farming activity carried out by individuals who reported farm income in the reference period ending in December 1999, 17 months prior to Census Day.

The results of our research have indicated the benefits of using tax data to maintain the coverage of the Farm Register. These benefits could be further exploited on several fronts. For the upcoming Census, we will investigate the use of tax data for reference year 2000, available only in December 2001, to augment the survey population used to assess the coverage of the Census with approximately 2,000 fiscal farms. The main factors to consider are the additional response burden put on farms that will have been correctly enumerated in the Census, and the quality of the response when contacting potential farm operators 8 to 12 months after Census Day.

We will investigate the feasibility of conducting a Farm Update Survey on an annual basis to maintain the coverage of the Register on a regular basis and, thus, maintain the coverage of the intercensal survey program and prepare for the 2006 Census.

Finally, we will conduct further research on applying the Farm Update Methodology to the population of between 25,000 and 30,000 corporate tax records that report farm income every year.

References

Statistics Canada (1997) 1996 Census of Canada: Agricultural profile of Canada, Catalogue: 93 356 XPB

Statistics Canada (1984) 1981 Census of Canada - Agriculture: Evaluation of Data Quality, Catalogue: 96 918

Lim A., Miller M., Morabito J. (2001) Research into improving frame coverage for agriculture surveys at Statistics Canada, *Proceedings of the Second International Conference on Establishment Surveys,* 131-136.

Estimation Using Estimated Coverage in a Census [(1)]

Wayne A. Fuller
Iowa State University, Department of Statistics and Statistical Laboratory,
Ames, Iowa 50010-1210, USA
e mail: waf@iastate.edu

Cary T. Isaki
U.S. Bureau of the Census, Washington, DC 20233-4200, USA
e mail: cary.t.isaki@census.gov

Abstract: Construction of population estimates using a census, or a sample from a census, and estimates of the coverage error in the census are outlined. A form of regression estimation based on quadratic programming is the basic estimation technique.

Keywords: Regression estimation, Census adjustment

1. Introduction

We consider the situation in which a subset of the population is observed and estimates of some population totals are available from another source. The subset might be a sample or it might be a census where it is known that coverage in the census is not complete. The characteristics of the population for which external estimates are available are called control characteristics or *X*-characteristics. The control estimates might come from a census coverage study or from a large national survey. For example an area sample could furnish estimates for a number of important crops. Because of coverage errors, the national survey may provide better estimates of crop acreages, or of farms with the crop, than the agricultural census. A third example is subsampling conducted as part of the census.
Our study was motivated by two situations associated with the 2000 U.S. Census of Population: the undercount and subsampling. It is known that the U.S. census has an overall undercount on the order of 1.2 percent. It is also known that the undercount varies by type of individual. For example, one of the higher rates is for young Hispanic male renters where the undercount may be as high as thirteen percent. A procedure to estimate the undercount was conducted as part of the 2000 U. S. Census and aggregate estimates of undercount were released in February 2001. (*U.S.A. Today*, February 15, 2001). Given the Census numbers and an estimate of the undercount, the objective is to produce estimates of the "census type" for person and household characteristics for small areas. A small area

(1) This research was partly supported by Cooperative Agreement 43-3AEU-3-80088 between Iowa State University, the National Agricultural Statistics Service and the U.S. Bureau of the Census. We thank Mingue Park for useful comments.

might be a small geographic area or it might be a subpopulation defined by characteristics of the population element.

In the U.S. Census of Population there are two types of forms used to collect information. The short form contains questions on basic personal and household demographic characteristics. The long form contains these basic questions, but also has questions on economic and social conditions. The long form is distributed to a sample of about one sixth of the dwelling units. The controls for long form estimation can come either from the Census or from the coverage study. Given the Census numbers, estimates of the undercount, and given the long form subsample, the objective is to produce estimates of long form characteristics.

It is assumed that there will be many users of the final data and that they represent the full range of statistical sophistication. Most will be accustomed to tabular data for the reporting units. A file that has the appearance of the traditional census data file based on reporting units but that yields estimates adjusted for coverage errors would be highly desirable for most users. A data file that can be tabulated as if it were a listing of the entire population is said to be a *transparent* census file. For the U.S. Census, the transparent file would contain a listing of housing units, persons within the unit, their reported data, and the block identification (geographic location) of the housing unit. Each unit in the transparent file has a weight of one. Alternatively, the data file may contain units with weights other than one. To give the estimates physical realism, the weights would be integers. It has been the practice of the U.S. Census Bureau to produce two files based on long form data, one file for households and one file for individuals. The files have been constructed using controls from the census and are composed of the long form data with integer weights.

We outline procedures that can be used to produce a user-friendly file. The inputs for the procedure are the data for the subset of the population and the control estimates. The creation of the data file requires several steps. The operations we describe begin with the information for the subset of the population, the estimated control totals and estimates of the variances of the control totals. Given this information, the next step is the creation of a set of weights for the subset that, when applied to the subset, produce estimates that are in agreement with the control totals, where "agreement" is to be defined. Post stratification, ratio estimation, raking ratio estimation, and regression estimation are procedures that have been used in similar situations. We will build upon the regression procedure.

2. Regression Estimation

Any characteristic can be related to the control variables through a population regression equation

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + e_i \ , \tag{2.1}$$

where y_i is the characteristic of interest for the i-th individual, $\mathbf{x}_i$ is the vector of control variables for the i-th individual, $\boldsymbol{\beta}$ is the population regression coefficient

$$\boldsymbol{\beta} = \mathbf{M}_{xx}^{-1}\mathbf{M}_{xy},$$

$$(\mathbf{M}_{xx}, \mathbf{M}_{xy}) = N^{-1}\left(\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i, \sum_{i=1}^{N}\mathbf{x}_i' y_i\right)$$

N is the population size, and e_i is defined by subtraction.

Let $\hat{X}_j, j = 1, 2, ..., m,$ denote the control estimate for characteristic j, and let $X_{sj}, j = 1, 2, ..., m,$ denote the estimate from the subset for category j. One regression estimator can be written as

$$\hat{Y} = Y_s + (\hat{\mathbf{X}} - \mathbf{X}_s)\hat{\boldsymbol{\beta}}, \tag{2.2}$$

where $\hat{\mathbf{X}} = (\hat{X}_1, \hat{X}_2, ..., \hat{X}_m)$ is the vector of control estimates, $\hat{\boldsymbol{\beta}}$ is the regression coefficient for the weighted regression of y_i on $\mathbf{x}_i$,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\mathbf{x}_i'\mathbf{x}_i w_{(0)i}\right)^{-1}\sum_{i=1}^{n} w_{(0)i}\mathbf{x}_i' y_i,$$

n is the number of reporting units in the subset, $w_{(0)i}$ is an initial weight, Y_s is the subset weighted total for the characteristic, and $\mathbf{X}_s$ is the subset weighted total for $\mathbf{X}$. For example, $w_{(0)i}$ might be the ratio of the coverage study estimate to the census estimate for a category of reporting units. If one were dealing with a probability sample, the initial weight would be the reciprocal of the sampling fraction. The regression estimator can also be written as the linear estimator

$$\hat{Y} = \sum_{i=1}^{n} w_i y_i, \tag{2.3}$$

where the i-th weight is

$$w_i = w_{(0)i} + (\hat{\mathbf{X}} - \mathbf{X}_s)\left(\sum_{i=1}^{n} w_{(0)i}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\mathbf{x}_i' w_{(0)i}.$$

One way to compute the regression weights of (2.3) is to minimize

$$\sum_{i=1}^{n}(w_i - w_{(0)i})^2 w_{(0)i}^{-1} \tag{2.4}$$

subject to

$$\sum_{i=1}^{n} w_i \mathbf{x}_i = \hat{\mathbf{X}} \quad . \tag{2.5}$$

We modify the basic regression procedure in several ways. The weights for reporting units should be nonnegative to guarantee that estimates of positive quantities are positive. Therefore a restriction

$$w_i \geq K \ , \tag{2.6}$$

where $K \geq 0$ is a fixed number, is imposed on the weights. Minimizing (2.4) subject to (2.5) and (2.6) is a classical quadratic programming problem and Husain (1969) suggested using quadratic programming to obtain the w_i.

A second modification of the regression weights is to relax some of the restrictions in (2.5). In the problems under discussion there are often a large number of controls available. That is, the dimension of the potential **x**-vector is very large. If one attempts to use a large vector, some of the weights may be very large or it will be impossible to satisfy the constraints (2.5). Therefore, in practice, some method of reducing the number of constraints, or of otherwise relaxing some of the constraints, is required. The most common practice is to reduce the number of constraints by collapsing categories.

Another procedure is to replace $m-L$ of the constraints in (2.5) with a component in the objective function. Then the quadratic program for weights for the mean for the large problem becomes; minimize

$$\sum_{i=1}^{n} (w_i - w_{(0)i})^2 w_{(0)i}^{-1} + \mathbf{CAC}' \tag{2.7}$$

subject to

$$\sum_{i=1}^{n} w_i x_{ji} - \hat{X}_j = 0, \qquad j = 1,2,...,L \ , \tag{2.8}$$

and subject to the bound (2.6), where

$$\mathbf{C} = (\sum_{i=1}^{n} w_i \mathbf{x}_{m-L,i} - \hat{\mathbf{X}}_{m-L}) \ ,$$

$$\mathbf{x}_{m-L,i} = (x_{L+1,i}, x_{L+2,i}, ..., x_{m,i}) \ ,$$

$$\hat{\mathbf{X}}_{m-L} = (\hat{X}_{L+1}, \hat{X}_{L+2}, ..., \hat{X}_m) \ ,$$

L is the number of exact constraints, and $\mathbf{A}$ is a fixed matrix. If we ignore the bound on the weights, the estimator can be written

$$\hat{Y} = Y_s + (\hat{\mathbf{X}} - \mathbf{X}_s)\,\boldsymbol{\Gamma}\,\hat{\boldsymbol{\beta}} , \tag{2.9}$$

where

$$\begin{aligned}
\boldsymbol{\Gamma} &= (\mathbf{G} + \mathbf{M}_{xwx})^{-1}\mathbf{M}_{xwx} , \\
\hat{\boldsymbol{\beta}} &= \mathbf{M}_{xwx}^{-1}\mathbf{M}_{xwy} , \\
(\mathbf{M}_{xwx}, \mathbf{M}_{xwy}) &= \Big(\sum_{i=1}^{n} \mathbf{x}_i' w_{(0)i}\mathbf{x}_i ,\ \sum_{i=1}^{n} \mathbf{x}_i' w_{(0)i} y_i\Big) , \\
\mathbf{G} &= \text{block diag}\,(\mathbf{0}_{L\times L},\ \mathbf{A}^{-1}) ,
\end{aligned}$$

and we assume $\mathbf{M}_{xwx}$ is block diagonal to match $\mathbf{G}$. It is possible to transform the $\mathbf{x}$-vector so that $V\{\hat{\mathbf{X}}\}$ and $V\{\mathbf{X}_s\}$ are diagonal matrices. If $\hat{X}_j$ is independent of X_{sj} , if both are unbiased for X_j , and if $V^{-1}\{\mathbf{X}_s\}$ is proportional to $V\{\hat{\boldsymbol{\beta}}\}$, an approximately optimal $\mathbf{A}$ matrix is diagonal with elements

$$a_{jj} = \overline{w}_{(0)}^{-1} \frac{R_j^2 V\{\hat{X}_j\}}{\left\{ R_j^2 V\{X_{sj}\} + n^{-1}(1-R^2)\,[V\{\hat{X}_j\} + V\{X_{sj}\}] \right\} M_{xwxjj}} , \tag{2.10}$$

where R_j is the correlation between x_j and y, R^2 is the squared multiple correlation, M_{xwxjj} is the j-th diagonal element of $\mathbf{M}_{xwxjj}$, and $\overline{w}_{(0)}$ is the mean of the $w_{(0)i}$. Using $R_j^2 = 1$, the estimator for the total of x_j is the best linear combination of $\hat{X}_j$ and X_{cj}. In practice a single set of w_i will be constructed for all y-variables. Also, in the census adjustment situation, X_s may be a biased estimator of the total. Thus α_j will be chosen on the basis of several criteria.

Regression estimators that attempt to bound the weights have been considered by Huang and Fuller (1978), Bethlehem and Keller (1987) and Deville and Särndal (1992). Estimators with relaxed constraints have been investigated by Husain (1969), Bardsley and Chambers (1984) and Rao and Singh (1997). See, also, Deville, Särndal and Sautory (1993). Quadratic programming has been used by Bankier et al. (1997) and Isaki et al. (1999, 2000a, 2000b).

Let Y be the unknown finite population total for the characteristic of interest, let $\mathbf{X}$ be the unknown vector of finite population totals for the x-variables, and let the representation (2.1) hold. The error in the regression estimator (2.9) can be written

$$\hat{Y} - Y = (\hat{\mathbf{X}} - \mathbf{X})\,\boldsymbol{\Gamma}\boldsymbol{\beta} + (\mathbf{X}_s - \mathbf{X})\,(\mathbf{I} - \boldsymbol{\Gamma})\,\boldsymbol{\beta} + (\hat{\mathbf{X}} - \mathbf{X}_s)\,\boldsymbol{\Gamma}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + e_s - e_\tau + O_p(n^{-1}) , \tag{2.11}$$

where $\mathbf{X}_s$ is the weighted total of the **x**-vectors for the subset, e_s is the weighted total of e_i for the subset, e_τ is the total of e_i for the population, and *n* is an index of sample size. In our examples *n* is the number of reporting units in the subset of the population. If $E\{\mathbf{X}_s\}$ is not **X** and if $\mathbf{\Gamma}$ is not **I**, then the estimator is biased. Thus in the case of coverage errors a portion of $\mathbf{\Gamma}$ will be the identity matrix.

Under the assumptions that the error in $\hat{\boldsymbol{\beta}}$ is independent of the error in $\hat{\mathbf{X}}$, that the covariance between $e_s - e_\tau$ and $\hat{\boldsymbol{\beta}}$ can be ignored, that $\mathbf{\Gamma} = \mathbf{I}$, and that the errors in $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{X}}$ decline at the rate $n^{-1/2}$, an estimator of the variance of $\hat{Y} - Y$ is

$$\hat{V}\{\hat{Y} - Y\} = (\hat{\mathbf{X}} - \mathbf{X}_s)\,\hat{\mathbf{V}}\{\hat{\boldsymbol{\beta}}\}\,(\hat{\mathbf{X}} - \mathbf{X}_s)' + \hat{\boldsymbol{\beta}}'\,\hat{\mathbf{V}}\{\hat{\mathbf{X}}\}\,\hat{\boldsymbol{\beta}} + \hat{V}\{e_s - e_\tau\}\ , \tag{2.12}$$

where $\hat{\mathbf{V}}\{\hat{\mathbf{X}}\}$ is the estimated covariance matrix of the vector of estimated control totals, and $\hat{\mathbf{V}}\{\hat{\boldsymbol{\beta}}\}$ is an estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

3. Long Form Estimation

To illustrate the procedures, we use data from the 1990 U.S. Census of Population. We constructed two sets of weights for the long form sample. One set used Census data as controls and the other set used Census coverage data as controls. The weighting of long form data is performed for geographic areas called weighting areas. We use WA1788, the largest weighting area in Houston, Texas. In 1990, the Census reported 8034 occupied housing units containing 25,145 persons in 176 Census blocks. The long form sample contained 2841 persons in 897 occupied housing units. Thus the realized average sampling rate is about one-in-nine.

The *x*-vector was not transformed and the Census totals were treated as fixed. When the Census data were used as controls, the objective function for the quadratic program was (2.7) with $L = 16$ equality constraints and $m - L = 67$ soft constraints. The matrix **A** was a diagonal matrix with elements that approximate (2.10) with $R^2 = 0.50$. There are five age, three race, one Hispanic, one sex, one renter, two household tenure, and three sub area variables in the equality set. There are 45 age-race-sex variables and 22 housing variables in the soft constraint set. Using such control variables means that the marginals of two-way tables agree with the controls, but totals for interior cells may deviate from the controls. We prefer this to a collapse procedure in which some marginals can differ from the controls.

The nominal sampling rates for the long form are one-in-two, one-in-six, and one-in-eight. The initial weights used in the quadratic program are the nominal sampling rates, denoted by $w_i^{(1)}$ modified to reflect the realized sampling rates. Cells were formed on the basis of demographic and geographic characteristics. The $w_i^{(1)}$ are then modified in a procedure that is a compromise between the direct ratio adjustment and a procedure that collapses cells with a small number of respondents. The modified weights are used as the initial weights $w_{(0)i}$ of (2.10). The lower bound K was set equal to one.

The elements of the diagonal matrix **A** were relatively large and, as a result, the estimates for the 67 soft constraints differed moderately from the controls. The largest difference was five and the largest percentage difference was four percent.

The average of the squared weights was 88, which is about ten percent larger than the square of the average weight. The sum of weighted squares in (2.7) divided by the number of households was 0.466. The smallest weight was equal to the bound of one and the largest weight was 30.

A second set of weights was constructed using person controls from the 1990 Post Enumerative Survey. The Post Enumerative Survey produces no household estimates. We used five age groups, four race categories, Hispanic, renter, and sex to form eleven equality controls. Sixty-three cells formed from the classifications were used as soft controls. An example of a cell is non Hispanic black males age 18 to 64 who live in an owned residence.

The estimated total number of occupied housing units was 8177 using the person controls from the Post Enumerative Survey, representing an estimated 1.8 percent undercount of housing units. The average of the squared weights was 92, the smallest weight was one, the largest weight was 29, and the sum of weighted squares divided by the number of housing units was 0.52.

The last step in weight construction was to round the weights to integers. Sample housing units were grouped by race/ethnicity of the householder and by tenure. Within each group the sample was sorted by family type by household size. Then the weights were rounded using the cumulate-and-round procedure.

The variances of estimators were estimated using replication. A random fraction of the sample was deleted. Then the quadratic programming weights were computed for the remaining part of the sample. In order for the variation among the replicates to correctly reflect total variation, when the controls are from the Past Enumerative Survey, the control totals for the replicates varied from replicate to replicate in a way such that the estimated variance of a control total using the replicates is equal to the directly estimated variance. See Fuller (1998).

References

Bankier, M., Houle, A., and Luc, M. (1997) Calibration Estimation in the 1991 and 1996 Canadian Censuses. Statistics Canada, (Draft), 8 pages.

Bardsley, P. and Chambers, R. L. (1984) Multipurpose estimation from unbalanced samples. *Applied Statistics*, 33, 290–299.

Bethlehem, J. G. and Keller, W. A. (1987) Linear weighting of sample survey data. *Journal of Official Statistics,* 3, 141–153.

Deville, J. and Särndal, C. (1992) Calibration estimators in survey sampling. *Journal of American Statistical Association*, 87, 376–382.

Deville, J. C., Särndal, C. E., and Sautory, O. (1993) Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013–1020.

Fuller, W. A. (1998) Replication variance estimation for two-phase samples. *Statistica Sinica*, 8, 1153–1164.

Huang, E. T. and Fuller, W. A. (1978) Nonnegative regression estimation for survey data. Proceedings of the Social Statistics Section, American Statistical Association 1978, 300–303.

Husain, M. (1969) Construction of regression weights for estimation in sample surveys. Unpublished M. S. thesis, Iowa State University, Ames, Iowa.

Isaki, C. T., Tsay, J. H., and Fuller, W. A. (1999) Weighting sample data subject to independent controls, Proceedings of the Statistical Society of Canada, Survey Methods Section, 21–25.

Isaki, C. T., Tsay, J. H., and Fuller, W. A. (2000a) Estimation of Census Adjustment Factors, *Survey Methodology*, 26, 31–42.

Isaki, C. T., Ikeda, M. M., Tsay, J. H., and Fuller, W. A. (2000b) An Estimation File that Incorporates Auxiliary Information, *Journal of Official Statistics*, 16, 155–172.

Rao, J. N. K. and Singh, A. C. (1997) A ridge-shrinkage method for range-restricted weight calibration in survey sampling. Proceedings of the section on survey research methods, *American Statistical Association*, 57–65.

Quality in Agricultural Statistics

Ulf Jorner
Statistics Sweden, 23 Klostergatan, SE-701 89, Örebro, Sweden
e-mail: ulf.jorner@scb.se

Fred Vogel
National Agricultural Statistics Service, 1400 Independence Avenue, 20250 2000
Washington, D.C., USA, Tel: 202 6908141
e-mail: fvogel@nass.usda.gov

Abstract: In recent years, the concept of quality in statistical data has developed in agricultural as well as in other statistics. One development is towards a broader definition of quality, including aspects as availability, clarity, confidentiality and security. Another is the trend towards user-defined quality, giving users a more active role in the process of weighing one quality component against another.
Quality in agricultural statistics also has an influence on the process that produces the statistical data. Quality assurance is not as much concerned with the individual data sets as with the agency's procedures.
The final measure of the quality of statistical data is whether users believe them. To achieve this, agencies must work both internally with their processes and externally with their users, while keeping a high integrity.

Keywords: Quality components, User-defined quality, Metadata, Quality assurance, Confidentiality, Security.

1. Introduction

There is a long history of discussions about quality issues related to the collection of data and the preparation of statistics. Deming (1944) was one of the first to point out issues concerning errors in surveys. The Hanson, Hurwitz, and Bershad (1961) paper on measurement errors in censuses and surveys was one of the first to make the point that errors affecting surveys were also a census issue.
Governmental organizations that produce official statistics have always been concerned with the quality of the statistics they produced and developed measurements such as accuracy, timeliness and relevancy. Accuracy was generally defined in terms of sampling errors, or total survey error. The main measure of timeliness was the period from the reference point to data dissemination. Relevance was more difficult to define, but it mainly meant whether the data helped answer the questions of the day. The main point is that the first measures of quality were defined by the producing organization. Some examples of quality measures are described in Vogel and Tortora (1987).

There is no generally accepted definition of the concept of quality, but a typical modern definition would be something like "the ability of a product or service to fulfill the expectations of the user". From this starting point, Statistics Sweden in 2000 adopted the following definition

> Quality of statistics refers to all aspects of statistics that are of relevance to how well it meets users' needs and expectations of statistical information.

Obviously, this and similar definitions need to be operationalized in order to be useful; we will return below with examples of how this is done in Sweden and in the USA.

One main point is that the modern approach is user-oriented, while the earliest measures of statistical quality were producer-oriented. Another point is that it is relative rather than absolute and even to some extent subjective. At least, different users tend to value a given component in different ways. As an example, a research worker would rank accuracy high, while a decision-maker would put more emphasis on timeliness. A final point is what is not included, i.e. costs. While costs are sometimes included in the quality concept, thus making a low cost contribute towards a higher quality, this has distinct disadvantages. The most obvious is perhaps that in general usage, quality is often weighed against costs. It is good to keep this possibility, so that also for statistics we may speak of value for money.

The purpose of this paper is to describe how the measures of quality need to change to become more customer focused. The rapid development of the Internet technology, for example, means that an organization's products are available around the world and used for purposes far beyond the original intent. The rapid access to data on the web is changing the meaning of timeliness and the increased sophistication of data users is also raising the need for new measures of quality. The authors will describe how the National Agricultural Statistics (NASS) in USA and Statistics Sweden are attempting to redefine how they monitor and report the quality of their products. At the same time, it should be stressed that it is an evolution, not a revolution. Original measures of quality, such as accuracy, are not superseded by new components, but rather supplemented.

A unique nature of agricultural statistics is that many of the items being measured are perishable or have a relatively short storage time. Many of the products are seasonable rather than being available on a continuous flow. It is also generally the case that the prices are highly volatile. Small changes in quantities translate into much larger changes in price. This places a premium on accurate forecasts and estimates. Because of the perishable and seasonable nature of agricultural products, the most important agricultural statistics are those that can be used to forecast future supplies. Knowledge of crop acres planted, for example, if known shortly after planting provide an early season measure of the future production. Inventories of livestock breeding stock can be used to forecast future meat supplies.

2. Changing Concepts of Quality

The most significant event is that the concepts of quality are being customer or externally driven rather than determined internally by the statistical organization. The availability of official statistics on the Internet is increasing the audience of data users. In previous times, the data users were fewer in number and usually very familiar from long experience with the data they were being provided. The new audience of data users is increasingly becoming more sophisticated and also more vocal about their needs.

2.1 The American example

The following paragraphs describe the six components of quality used by the National Agricultural Statistics Service and how they are measured:

- Comprehensiveness
- Accuracy
- Timeliness
- Comparability
- Availability
- Confidentiality and security

1. Comprehensiveness: The measure of quality is whether the subject matter is covered. In agriculture, one example is that crop production is provided by including estimates of acres planted and harvested, yields per acre, production, quantities in storage at the end of harvest, measures of utilization such as millings, exports, feed use, and the average prices to measure value of production. A comprehensive view of crop production includes the land use, the yield that is a measure of productivity, total production and stocks that measure the total supply, and the utilization that measures the disappearance. One way the NASS evaluates the comprehensiveness of its statistics program is through annual meetings with organizations and people using the data. For example, every October, the NASS sponsors a widely advertised open forum and invites all interested people to meet with subject matter specialists to seek their input about the content, scope, coverage, accuracy, and timeliness of USDA statistics. Input received at recent meetings was that because of the emergence of bio tech crop varieties, information was needed about the proportion of the crop acreages planted to those varieties. As a result, the data on acres planted and for harvest for several crops are now provided by bio tech class.
NASS also has a formal Statistics Advisory Committee comprised of producers, agribusinesses, academia, and other governmental users. This advisory committee meets once a year with NASS to review the content of the statistics program and to recommend where changes in agriculture are requiring a change in the NASS program. It is a result of these meetings that NASS will be including questions about contract production in agriculture in the next census of agriculture.

2. Accuracy: Traditional measures include sampling errors and measures of coverage. These are basic measures that need to be used by the statistical organization in its internal review of its data. Certainly, the organization needs to define standards for sampling errors and coverage that determine whether or not to publish data not meeting those standards. The NASS also publishes a measure of error for major data series and uses a method called the root mean square error which is the difference between an estimator and the “true”

value. The root mean square error (RMSE[1]) as defined by NASS for this usage is the average difference between the first estimate or forecast and the final estimate. Using corn as an example, NASS publishes a forecast of production in August each year. This forecast is updated each month during the growing season. After harvest, a preliminary final estimate of production is published. A year later, the preliminary final estimate may be revised based upon utilization and stocks data. Each five years after the census of agriculture, the estimates of production for the census year and years between census periods are again reviewed and revised if suggested by the census results. The root mean square error is published each month during the growing season and each time a revision is published. The reliability table also shows the number of times the forecast is above or below the final and the maximum and minimum differences. The primary gauge of the accuracy of reports is whether they meet the needs of the data users. If revisions are below their expectations, NASS will be pressed to either improve the accuracy, or discontinue the report.

3. Timeliness: One measure of timeliness is the time between the survey reference period and the date of publication. There is a trade-off between accuracy and timeliness. Timeliness needs vary depending upon the frequency the data are made available. Those using data to make current marketing decisions need the data to be current. It is not helpful to not learn until long after harvest and after much of the crop was already sold that the current year's corn crop was at a record high level. On the other hand, someone making long-term investment decisions may be more concerned whether the census results are consistent with history. NASS has the following standards for timeliness. Monthly reports with a first of month reference date should be published during the same month. The report should be published before the next data collection period begins. Quarterly and semi-annual reports should also be published within one month of the reference dates. The census data published every 5 years should be published one year and one month after the census reference period. The 2002 census will begin on January 1, 2003. All revisions to the 2002 crop year and January 1, 2003 preliminary estimates will be published by early January 2004. The target to publish the census results will then be February 1.
Another issue of timeliness is that the data user should know when the data are to become available. Each October, NASS publishes a calendar showing the date and hour that over 400 reports will be published during the coming calendar year. The final issue related to timeliness is the need to be punctual and meet the dates and hours as published in the calendar.
Market sensitive reports should not be released while the markets are open. There should be a policy to only release such reports in the morning before markets open or at the end of the day when markets have closed. This may became a moot point some day when electronic

---

$$RMSE = \frac{(\Sigma(\aleph\iota - X\iota)^2)^{1/2}}{n} \text{ where}$$

[1] $\aleph\iota =$ first estimate or forecast

$X\iota =$ final revised estimate

trading allows continuous marketing, but until then, the timing of the release of the data needs to be considered.

4. Comparability: Data should be comparable over time and space. It is necessary to carefully define for the data users what is being measured by defining the basic characteristics and ensure they remain the same. If it becomes necessary to change a concept or definition, then there should be a bridge showing how the data relationship changed. This may require that a parallel survey be conducted for a period to measure the affect of the change in definition. Comparability is essential for measures of change. Even subtle changes such as the way a question is asked on a survey form can change the final result.
Data should be comparable to other related information or there should be an explanation of the reasons for the departure. An example would be when livestock inventories are declining, but meat supplies are increasing. The reason could be that animals are being fattened to much heavier weights. The statistical organization needs to "mine" its data to ensure it is comparable to internal and external sources, or to explain the difference.
Data should be easily understood or detected. NASS recently published some data showing the percent of the corn acreage planted resistant to Bio-tech herbicide and the percent planted resistant to insect varieties. The percents were not additive because some seeds contained both characteristics. However, data users added them anyway because the arrangement of the data made them think they could. The point is that statistical organizations need to ensure that appropriate inferences can be made about published data even if it requires additional tables or explanation. If pieces are not additive, or totals are not the same as the sum of the parts, then they must be presented so that there is clear understanding.

5. Availability: Official statistics need to be available to everyone and at the same time. Data collected using public funds should be made available to the public. The integrity of the statistical agency should not be at stake by granting special access to some people or organizations before others. Special care needs to be taken to ensure that there is no premature release of information - all users need to be treated equally and fairly. NASS operates under laws and regulations that require strict security measures ensuring that data are released only at the appropriate time and place. In addition, the data should be made available immediately to everyone via printed copies, press releases, and the web site.

6. Confidentiality and Security: The quality of official statistics is dependent upon maintaining the trust of the farms and agri-businesses reporting for their operations. A basic tenant should be that their data are strictly confidential and used only for statistical purposes. The statistical agency should seek laws and regulations that provide protections from the data being used for taxation and/or regulatory purposes. As the data are tabulated and preliminary estimates being derived, there need to be precautions to prevent premature disclosure of the results. First, a premature disclosure may be misleading if it differs from the final results. It may also give some data users an unfair advantage in the use of the data. The key to an organization maintaining its integrity and public trust is to first being totally committed to maintaining confidentiality of individually reported data and ensuring the

security of the estimates until they are released to the public. None of the data in the 400+ reports NASS publishes each year are subject to political approval. In fact, they are not even made available to cabinet officials including the Secretary of Agriculture until they are released to the public. In some instances, the Secretary will enter NASS secure work areas minutes before data release to obtain a briefing about the report.

2.2 The Swedish Example

The Swedish agricultural statistical system is, apart from scale, rather similar to the American system. The same can be said about the trends in quality concepts. The Swedish experiences will thus not be given as duplicates, or for a comparison. Rather, they will be used to expand the discussion of changing concepts of quality.

The Swedish quality concept for official statistics has five components:

- Contents
- Accuracy
- Timeliness
- Coherence, especially Comparability
- Availability and Clarity

For a background to this, as well as some useful references, see Elvers and Rosén (1999).

1. <u>Contents</u>

This component is very similar to the Comprehensiveness component of NASS. It is interesting to note that this type of component was earlier often called Relevance. The reason behind the change is that what is relevant for one user may be irrelevant to another. A good example is given by the use of lower thresholds. In Swedish agricultural statistics, this threshold is 2 hectares of arable land (or large numbers of animals or significant horticultural production). This reduction of the population of farms will rule out less than 1% of crop production and would thus be acceptable to most users. On the other hand, it may be very disturbing for persons studying rural development in less favored areas. Thus, with a user-oriented quality approach it seems more natural to talk about contents or comprehensiveness.

An interesting aspect of Swedish statistics is the great use made of administrative data. This source of data is also prominent in agricultural statistics. As a prime example, the Common Agricultural Policy in the European Union makes a wide range of subsidies available to farmers. For most crops and animals, the applications from farmers give a very good coverage of the population in question. This forms a very cheap source of data, and moreover the data have been checked extensively, without cost to the statisticians. On the other hand, the statisticians have no control over the definitions used, the thresholds applied, the crops included or excluded, etc. Much work is thus required to convert administrative data into statistical data. In Sweden, the definitive way to do this has not yet been decided upon, but Wallgren and Wallgren (1998) gives a good overview of

possibilities and problems. This issue is also the subject of a paper by Widén and Jönrup in this volume.
With administrative data, another problem is that the source may run dry due to political or administrative decisions. As an example, tax returns were for many years used in Sweden to produce statistics on farmers' income (with due regard for possible systematic errors). A simplification of Swedish tax legislation made farm incomes virtually indistinguishable from other forms of income in tax returns, and thus this statistical series was ended.

2. Accuracy
When measures of accuracy became available in the mid fifties, this measure tended to become almost synonymous with quality. Lately, it has dropped in importance compared with e.g. timeliness and availability.
Accuracy of course involves both precision and bias. While the former is easily measured, e.g. as standard error, the latter is both difficult and expensive to measure. As an example, reporting errors in crop areas were measured in the field for a random sample of farms in Sweden up to 1995, but was the discontinued for cost reasons. Even when measurements were made, the number of measurements did not allow error estimates for individual years, but rather estimates for a time period. In a similar way, nation-wide coverage checks were never made for the Swedish Farm Register, although spot checks indicated that the undercoverage of crop areas and livestock was negligible while the number of farms may be underestimated by 1-2%.

3. Timeliness
Timeliness is perhaps the component of quality that has gained most in importance by the transition from producer-oriented to user-oriented quality. It is also naturally in conflict with components as accuracy, as shorter production time means less time to follow up possible sources of errors. One attempt to both have the cake and eat it is to use preliminary statistics. However, there is some price to pay with regard to clarity, as in the American example above.
While timeliness is one of the best examples of components that have received high priority lately, it is interesting to note that as early as 1877 it was one of three components (expedition, completeness and reliability) that defined quality in Swedish agricultural statistics.

4. Coherence, especially Comparability
The more explicit role of users in the process of choosing appropriate levels of quality also highlights the eternal dilemma between the needs for changes in definitions, etc. and the need for long time series. Users like research workers will put priority on long, undisturbed time series and thus good comparability over time, while policy-makers are more interested in statistics that give a good and up-to-date picture of current agriculture. Official statisticians often have to devise compromises; a good but often expensive solution is to provide parallel series for a few years at the point of change. As an example, when the population and domains of study were changed for the Swedish Farm Accountancy Survey at Sweden's entrance in the European Union, series for both the old and new classification have been published for both 1997 and 1998.

5. Availability and Clarity

Statistics Sweden makes annual assessment of changes in the five quality components, and the one with the highest level of improvement over the last years is Availability and Clarity. This holds true also for agricultural statistics.

The increased availability of agricultural statistics makes metadata more and more essential. Metadata is data about data; the term was coined as early as in 1973, cf. Sundgren (1975). As a simple example, adopted from Sundgren, the figure 427 621 is quite meaningless as such; the information – metadata -that it is the number of milk cows in Sweden last year gives it some meaning. However, serious users of statistics would need more background information, e.g. what date it refers to (2 June 2000), if it is based on a sample (yes, n = 10 009), the data collection method (mail questionnaire), etc. Other users would be interested in non-response rate (6.3%), standard error (0.6%), or in the properties of the frame. Still other would like to know how this figure could be compared to earlier figures for Sweden (based on different techniques) or to figures for other countries, or if there is a regional breakdown available (and if so, the corresponding quality measures).

Two aspects of metadata should be stressed here. First, different users have different need for metadata, thus the system to retrieve it must be flexible, and of course preferably easily accessible via e.g. Internet. Second, metadata should not only act as some kind of tag on the actual figures, it should also provide a means to search for information. A user should be able to search for e.g. 'milk cows' and find the appropriate statistical series and its properties. Good metadata is thus both an end and a mean as regards good availability.

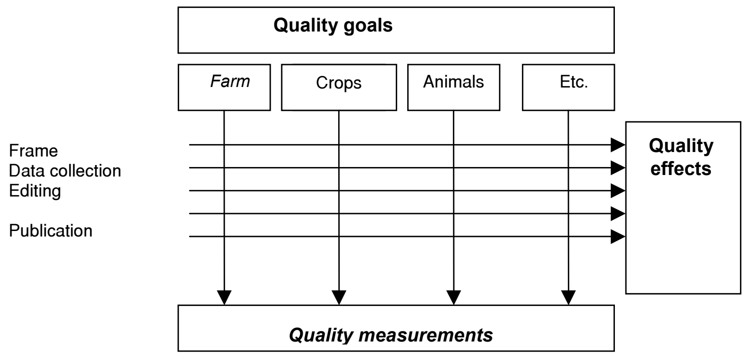
3. Assuring Quality

Measuring quality, however difficult it might be in specific cases, is only a lower level of quality ambition. Setting and meeting quality goals are, both from the user's and the producer's perspective, a higher ambition. The term quality assurance is often used for the process of ensuring that quality goals are consistently met.

3.1 Quality Assurance as an Agency Undertaking

Traditionally, we tend to think of quality as a property of a single variable. We measure this quality and try to control and hopefully improve it. Of course, we realize that the quality properties of different variables are inter-connected, as they are often collected in the same survey, or that they share common processes. As an example, the same frame is normally used for different surveys, and the same field staff, etc. Thus, we must consider a whole system when considering quality assurance. The figure below has been adopted from Medin et al. (1991) to illustrate this point.

Figure 1. *The system of quality assurance*



The figure illustrates not only the fact that the variables are interconnected, but also that we generally have problems to measure the effects of individual efforts to improve quality; measurements are made in one dimension whereas effect appear in another. To assure quality at each intersection of the grid is impractical, thus quality assurance should apply to the whole system producing the agricultural statistics in question, or more appropriately even to the whole agency that produces it.

It is useful to separate the quality of the statistics from the quality of the process that produces it. Thus, a process that assures a pre-determined level of quality in a cost-efficient way is in itself "good", even if the quality level is low. In a way, quality assurance can be looked upon as the link between quality in the process and quality in the product.

Most agencies producing agricultural statistics have been subject to budget cuts, but have been able to reduce the adverse effects on the quality of agricultural statistics through improvements in processes. The application of new techniques, as computers and the Internet, as well as new data sources, e.g. administrative registers and remote sensing, have been part of this process, but also new management ideas, as e.g. Total Quality Management (TQM) and Business Process Reengineering (BPR). It should be remembered that an approach as TQM involves both gradual improvement in existing procedures and development of totally new procedures.

Finally, it should be stressed that maintaining the integrity of an agency, e.g. by having and demonstrating independence from the political sphere, is of paramount importance in assuring quality in the eyes of users.

3.2 Examples of Quality Assurance Efforts

The statistical agency should adhere to all of the traditional principles of quality control that have been well documented elsewhere. The entire statistical process from frame development needs controls imbedded to control quality. Basic things such as data edits and analysis are crucial and should be a normal part of doing business. The following

paragraphs will describe some non-traditional things that NASS has been doing to enhance data quality.
Technical Review teams: Teams have been formed for two purposes. NASS has an office in 45 states in addition to the headquarters units. The headquarters units develop the specifications for the survey, analysis, and estimation processes that are carried out by the state offices. To ensure that the states are properly interpreting and carrying out the correct procedures, teams of subject matter people are periodically sent to visit state offices on a rotating basis to do a complete review of the office. The team is responsible for documenting their findings, and the state is responsible for correcting areas that did not meet the standards. It is sometimes found that one reason the states were not following correct procedures was because the headquarters' instructions were not clear or complete. The other purpose of technical teams is to review a specific estimating program. The most recent effort involved a review of the survey and estimating processes used to generate estimates of grain in storage. A team of experts reviewed the statistical and data collection processes from the state offices through the preparation of the final estimates. As a result of this review, the sample design is being revamped, and a new edit and analysis system is being developed.

Data warehouse: NASS has developed a data warehouse system that provides the capability to store all data for every farm in the U.S. It currently contains individual farm data for the 1997 census of agriculture and all survey data collected for each farm since that time. Edit and analysis systems are being developed that will incorporate the use of historic information on an individual farm basis.

Statistical Research: The Statistical research division in NASS operates outside the operational program. It does periodic quality check re-interviews, special analysis to identify non sampling errors, develops new sample designs and estimators, and provides in-house consulting services to monitor and improve quality of statistics.

Hotline: A staff is assigned to responding to a toll free telephone number or an email address. Any farmer or data user in the country can make a toll free call to this number or send an email message to ask any question or raise any concern about anything NASS does. A log is maintained and used to determine areas needing improvement.

Outreach: As described above, NASS relies upon an advisory Board for input in addition to the formal program of data user meetings. NASS staff also meet regularly with commodity organizations and other governmental organizations to keep current with their needs for quality.

4. Conclusions

As data users become more sophisticated in their use of public data, their demands for quality will be a major force to determine the quality of official statistics. Data producers must take into account that data users will attempt to use information for purposes that go beyond the capabilities of the data. As an example, user-friendly quality declarations/metadata will be very important.

Quality assurance of agricultural statistics will also be more important, in order to prevent large accidental errors that might lead users to wrong and perhaps costly decisions. Such assurance will apply to the agency as such rather than to individual statistical series. Use of unconventional approaches and new techniques will be ever more important in this work.

The trends in USA and Sweden are probably representative for other countries as well. Of course, there are deviations from country to country, e.g. in how far these trends have progressed, and the relative importance of different quality components. Hopefully, the guiding principles defining quality in Sweden and the U.S., could provide examples and standards that will be useful to other countries.

The final measure of quality is whether the statistics being produced are believable. The ultimate test is whether the public accepts the data as being accurate assessments of the current situation. This public perception is ultimately based on its trust in the integrity of the statistical organization. The foundation of any statistical organization's integrity is to have the public's trust that it protects the confidentiality of individual reported data, ensures statistics are secure before release, and releases results to everyone at the same time.

References

Elvers E., Rosén B. (1999) Quality concepts for official statistics, in: *Encyclopedia of Statistical Sciences. Update Volume 3*. Kotz, S., Campell, B. & Banks, D (Eds), Wiley, 621-629.

Medin, K. et al. (1991) *Quality control system in the farm register* (in Swedish), Statistics Sweden, Stockholm

Sundgren B. (1975) *Theory of Data Bases*, Petrocelli/Charter, New York

Wallgren A., Wallgren B. (1998) *How Should We Use IACS Data*, Statistics Sweden, Örebro

Vogel, F. and Tortora (1987) *Standards for Official Statistics.* ISI Bulletin

6 June 2001

Invited Paper Session

AGRICULTURAL STATISTICS SYSTEMS

Chair: A. De Michelis

The Future European Agricultural Statistics System

Christian Gay
SCEES, Ministère de l'agriculture et de la pêche, 251 rue de Vaugirard, 75732 PARIS Cedex 15, France
e-mail: christian.gay@agriculture.gouv.fr

Abstract : In view of the increasing association of non-farm activities with the more traditional agricultural occupations, the development of the environmental concern, and the enlargement of the European Union, the European agricultural statistics system will have to undergo thorough changes.
The bottom-up system in which the last contribution was expected to produce the European estimate will not survive the enlargement, and top-down system directly producing a European figure that Member States could break down using additional surveys or administrative sources, must be contemplated.

Keywords : European Enlargement, Common Agricultural Policy, Statistical Methods, Multi-functionality, Administrative Statistics, Representativeness.

1. History

The present European agricultural statistics system has developed as the common market organisations were implemented. These statistics were a support to the rise of the common market organisations. For instance, three pig survey have been carried out each year after the setting-up of the pig market organization.
Since the CAP was created, common market organisations and specific statistics have developed at the same time. Each Member State was in charge of its own data, and the European data was the result of an aggregation of national data, in a bottom-up process.
This necessitated a similar periodicity in all States, and, for instance, to harmonize and synchronize various censuses and samplings was created as the Farm Structure Survey, carried out every year ending by 3, 5 or 7 between two censuses.
Now, although it has sometimes been argued that agricultural statistics get too much coverage when compared to the farming sector's contribution to GNP, it must be remembered that agriculture is the most administrated sector, and the effort put into agricultural statistics must be compared to the cost of public intervention in the agrifood sector.

2. Why a new system?

2.1 Internal enlargement

Traditional needs in terms of agricultural statistics: farm structure surveys, livestock surveys, production statistics, economic accounts, price statistics ... will continue to play a key role in years to come.

However, it is not possible any longer to restrict the data collection to the farming activity.

The diversification of the activity of farmers and the acknowledgement of the multifunctional role of agriculture in a broader approach of rural development will increase the need for information on activities related to agriculture such as processing and marketing of farm products on farms and farm tourism.

It seems that, since the Future Agricultural Data Outline seminar (FADO, 1998) the point of view of the statisticians has changed. In view of the increasing association of non-farm activities with the more traditional agricultural occupations, FADO focused on the integration of information on agricultural structure with a broader picture of economy. Nevertheless, the statistical of agriculture and its relation to the conservation of environment and consumer health was only at its beginning, although a territorial approach was mentioned several times.

Now multi-functionality and environmental concern impose themselves as a more prominent part of the agricultural policy.

In addition to its traditional agricultural products, the agricultural sector has important outputs in terms of environment, tourism and leisure services as well as the integration of processing and packaging activities on the farm.

The statistical system must be able to assess the qualitative improvement of agricultural production : monitor the development of sustainable agriculture, the reduction of the agricultural pollution potential, and the dissemination of more environment-friendly methods such as organic farming, as well as the production of non-food products such as biomass for renewable energy.

More specifically, the agricultural statistician must be able to assess the contribution of the agricultural activity to the preservation of the environment, as well as the risks resulting from inappropriate farming practices This means establishing agri-environmental indicators such as the use of pesticides and fertilizers, the spreading of waste and the changes in land use, land cover studies and evaluation of the landscape and the biological diversity.

Considering the emergence of these new topics, it will be more and more difficult to rely on the current bottom-up aggregation system in which, to produce the European estimate for a crop, awaits the last contribution, or procedures of replacement of the last missing data must be implemented.

Fortunately, at the same time as the multi-functionality of farming develops, assessment and control system develop too, and provide a large bulk of administrative data in which the statistician must be able to extract indicators.

We are thus discussing two readjustments:

- between European operations and national operations;
- between the respective use of survey statistics and administrative statistics.

This means that there is an absolute necessity to switch, partly and progressively, from aggregated statistics to surveys on the global European economic area.

In other terms, these surveys would be carried out with the objective of representativeness at the level of the European Union. Member States, and particularly big producers of a particular category of crops would be free to complement the samplings and the contents of the questionnaires to obtain the information that match their needs.

In this top-down system, complementing the European samples is only one way. Another way is to use the administrative by-products of the management of the agricultural policy.

The RICA-FADN (Farm Accountancy Data Network) may be considered as a kind of precursor of top-down systems, even if the sample is not designed to be representative at European level only, but allows comparisons between European regions.

A new one is currently being set up, the LUCAS (Land Use/Cover statistical Area frame Survey). This new European survey will, for the first time, provide harmonized EU statistics on land use and land cover, including non-agricultural uses and environmental information. LUCAS is not the first area frame survey to be conducted in Europe, but it is the first with a European coverage and representativeness, and is more modern than all previous operations, for its conception integrates the most recent working techniques of geographic information management.

2.2 External enlargement

At the same time, it has to make these readjustments: joint use of survey statistics and administrative statistics and setting up of a programme of surveys aiming at European representativeness, the EU will have to cope with the arrival of countries which are reaching market economy.

The incorporation of the systems of agricultural statistics of these countries into the European statistical system poses a real challenge.

In these countries, agricultural area, GNP contribution and employment in agriculture are more significant than the present EU average, and agriculture will thus play a key role in the enlargement negotiations.

Not only will it be necessary to ask these countries to integrate the " acquis communautaire ", but this confirms the need of European statistics itself to undergo a thorough reform.

The bottom-up aggregation system has already shown its weaknesses in the EU 15, will not be able to go on with the next waves of enlargement. With 25 or 30 Member States, it could be the end of a good agricultural statistics system in Europe.

There is another problem. While all EU 15 countries have more or less the same idea about the meaning and the methods of statistics, in the candidate countries, for a long time, doing statistics has meant adding up figures on the administrative forms that were filled to show compliance with the state plans.

Needless to say that this type of administrative data has nothing in common with the administrative data that EU 15 countries are used to utilize.

While EU 15 countries use reliable farmer registers, there is no such a thing in the candidate countries, where farm structures are undergoing permanent changes. Land is redistributed, production units are broken up, then merged again into new units.

Another problem is that the land redistribution caused the emergence of a great number of subsistence farms which are more extended family gardens than agricultural holdings in the sense of European statistics.

One of the targets in the next few years is to have farm structure surveys conducted in candidate countries according to EU requirements. Due to the fact that the farmer lists are not kept up to date, the construction of a sampling frame will be difficult and multi-frame surveys combining area frames and list frames must be contemplated.

3. Outlines of a future system

With the development of top-down surveys, the official statistical bodies of the Member States will have a double function:
- acting as national statistical agencies to provide national and local decision-makers and the public with the information they require;
- acting as executive bodies of a European statistical body which would not be any longer an office of the Commission, but an agency at the disposal of the different institutions of the European Union.

Since the entry into force of the Amsterdam Treaty, documents concerning agricultural statistics are included in documents formally submitted to co-decision. The European Parliament is associated by means of such a co-decision procedure.

All these challenges have to be faced in a context of reduction of human and financial resources.

This means than even more than in the past, it will be necessary to select the actions to be carried out with the utmost care.

One way of doing this is rather to adopt a European view of the problem, using appropriate European samples instead of the costly and moderately efficient aggregation procedures.

A major issue underlyies. Shall we organize European statistics into a public, harmonized and organized system, as an activity pertaining to the prerogative of the Union, or shall we, for lack of human and financial resources, allow the development of unofficial systems, only to the benefit of those that would finance them?

The European and national powers must be able to monitor and assess the economic measures. To that purpose, they need a performing and independent tool, with the capacity of making statistical data available to all, decision-makers and public.

Only a true public service of statistics will have such a capacity.

Agricultural Statistics System Based on Surveys as Well as Administrative Sources

Marie-Louise Widén
Statistics Sweden, SE 701 89 Örebro, Sweden
e-mail: marielouise.widen@scb.se

Hans Jönrup
Swedish Board of Agriculture, SE 551 82 Jönköping, Sweden
e-mail: hans.jonrup@sjv.se

Keywords: Linkage variable, Metadata, Register-statistical methods, Register statistics, Statistical register, System of registers.

1. Introduction

”It is quite clear that the Member States find themselves in the paradoxical situation of having to face a number of budget cutbacks at the same time as providing users with an increasing volume of high-quality relevant information”. These words by Mr Y Franchet at the Seminar on the Use of Administrative Sources for Statistical Purposes (Eurostat, Luxembourg, 15-16 January 1997) point to the demand for more efficient statistical systems. An increased use of administrative sources is a way to reduce costs and response burdens when producing statistics and a way for more efficient and flexible use of existing data in order to meet new statistical demands.

Sometimes scepticism is an expressed against administrative data and it is declared that administrative sources should not be used for statistical purposes because coverage and measurement quality may be poor. Against this it must be emphasized that administrative data should not be used directly as they are, because, in order to meet the statistical needs, they must be processed by what is called *register-statistical methodological work*. Furthermore, administrative sources must be supplemented by register maintenance surveys, regular statistical surveys or other administrative sources if this is necessary to obtain the desired quality.

The structural statistics on agricultural holdings have previously been based solely on traditional censuses and sample surveys. New administrative sources, such as the IACS (Integrated Administrative and Control System, the administrative system for agricultural aid), and all the potential information about the agricultural sector within the whole register system, is recognised as offering new possibilities for agricultural statistics.

2. New Possibilities for Agricultural Statistics

“The political changes in the Common Agricultural Policy (CAP) widen the scope of agriculture… What is more important in the new CAP is that the importance of agriculture in preserving and maintaining both the environment and social activities in rural areas is recognised, i.e. rural development becomes a second pillar of the CAP. Agricultural production no longer needs to be the main occupation of the holding to make it an “agricultural holding”. Moreover, agriculture policy becomes also concerned with the alternative sources of income and employment of the farming families. Agricultural statistics have to adapt their scope to this broader framework in the most efficient and creative way. To get the complete picture of agricultural and rural activities, new techniques should be adopted, by using both existing non-statistical, but reliable, data sources and by implementing other survey techniques…”

This quotation is from the executive summary to the item “Wide Agriculture” which was discussed by the Agricultural Statistics Committee (Copenhagen 29 & 30 June 1999). The quotation illustrates that an increased demand for new kinds of agricultural statistics needs to be met. One way to satisfy this demand is to use the new possibilities offered by register-statistical methods. This includes both the systematic uses of administrative sources like IACS, which are specially designed for the agricultural sector, and general administrative sources, e.g. data from the taxation system.

3. Register Statistical Terms and Methodological Issues

In this paper a *register* is defined as a complete list of objects belonging to a defined object set. The objects in the register are identified by *identification variables*, which make it possible to update the register and to link it with other registers. A *statistical register* is defined as a data set with identifiers where the object set and variables correspond to the statistical problem.

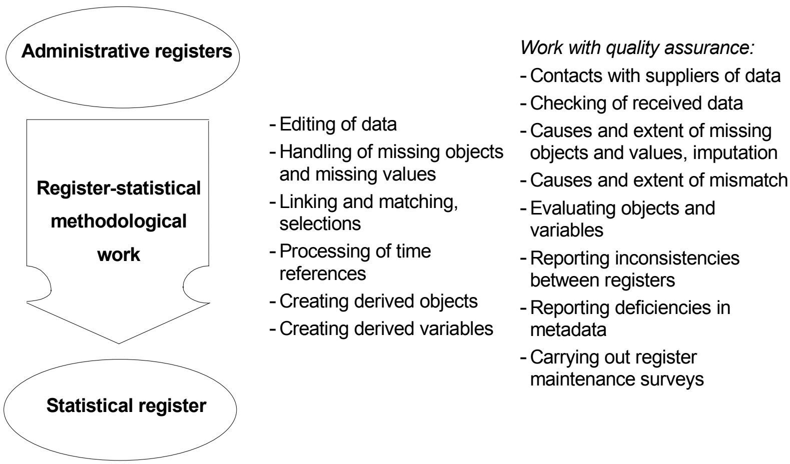
A *system of registers* consists of a number of registers that are linked to each other by one or more common identification variables or linkage variables. An efficient system requires that the linkage variables are of good quality and that the same linkage variables can be found in different registers. Furthermore the definitions of the objects and variables in the system must be harmonised so that data from different registers can be used together. Reference times must also be consistent.

Designing a register for statistical production is a different statistical task from designing a sample survey. In the case of *sample surveys* the first step is to determine the population and the parameters to be estimated. This in turn determines the character of the survey with regard to sample design and estimation. Thus the definitions of the population and parameters are determining the way of the data collection. As a rule *one* survey at a time is considered, with a limited number of parameters.

In the case of *register statistics* a different approach is taken, since some data have already been collected and are available in different registers that are not tailored for a statistical application. From available registers a selection is made of objects and variables that are relevant to the issue addressed by the statistical problem. In some cases, on the basis of available registers, new variables - and possibly new objects as well - have to be derived. Thus, the data are at hand first, and then they must be adjusted by register-statistical methodological work in order to be used when producing relevant statistics.

Metadata have a more important role when producing statistics from a system of registers than from sample surveys. Sample surveys can be documented one by one and usually have no connections with other surveys. Register-based statistics on the other hand may use data from different registers (in a statistical register system) which are based on administrative registers. For this kind of statistics there is therefore a great risk of errors if there is a lack of familiarity with problems of comparability and changes of definitions. Without knowledge about definitions and quality problems, the administrative registers cannot be used for statistical purposes. It is also important that metadata are tailored for register statistics.
So even though much of the statistical methodology is the same for survey statistics and register statistics - such as non-sampling errors and problems of analysis and presentation - the ways of thinking are different. Sampling errors and design problems are central to survey statistics while in register statistics the system approach is fundamental. In improving the quality you cannot look at one register at a time, you have to consider the system as a whole and pay special attention to identification variables used for linking purposes.

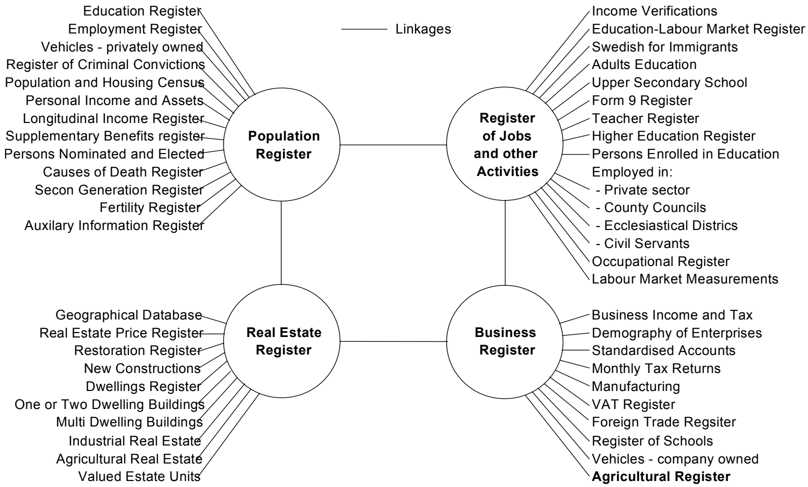
Figure 1: *Register-statistical methodological work on the register level*



4. Statistics Sweden's System of Statistical Registers

At Statistics Sweden there are many statistical registers based on administrative sources. The extent of the general register system can be seen from the figure below, which shows the most important registers. It shows the system Statistics Sweden wants to build, not the present system. Some parts do not exist today, new registers are under development and new linkages are planned; further, it is a hope that many of the present registers will be modified. By harmonization and more efficient use of the administrative sources the system will be made more efficient and the quality of the register statistics produced will be improved.

Figure 2: *Statistics Sweden's System of Statistical Registers structured by Type of Object*



A statistical register system is made up of the following parts:

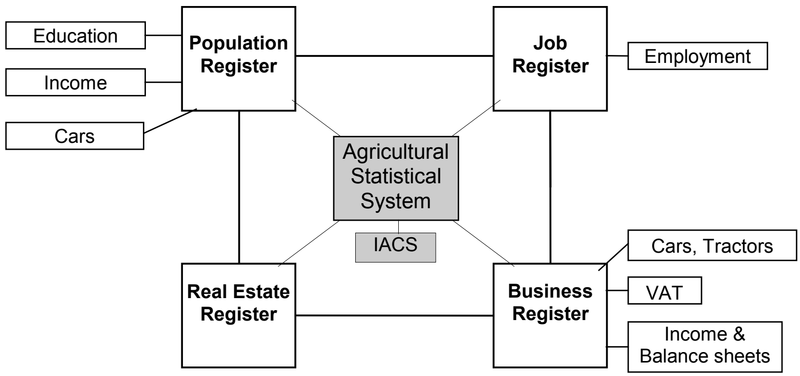
1. *Basic registers*: Statistical registers of *objects* of fundamental importance for the system. The four circles in the diagram are the basic registers.
2. *Other statistical registers*: registers with *statistical variables*. In the figure there are 48 other statistical registers with a large number of statistical variables describing the populations in the basic registers.
3. *Linkages* between the objects in different basic registers and between basic registers and statistical registers.
4. *Standardised variables*: variables of fundamental importance for the system.
5. *Metadata*: definitions of objects, object sets and statistical variables, information about quality and comparability over time should be easily accessible.
6. *Register-statistical methods* including routines for quality assurance.

In the Nordic countries population registers for statistical purposes are used and these registers are very important for the statistical systems. The laws regulating this practice guarantee individuals a high level of confidentiality. All register processing at Statistics Sweden involving statistics on an individual level is under the supervision of a governmental authority independent of Statistics Sweden.

5. An Agricultural Statistics System within the General System of Statistical Registers

In the future it will be possible to combine the Agricultural Statistics System with the Statistical Registers in the General System. By using data from different registers with complementary information from statistical surveys the possibilities are increasing to produce relevant statistics about agriculture and other activities (widening agriculture) with less burden on the farmers.
Such a system could be as follows:

Figure 3: *Agricultural Statistics system inside a general system*



6. Agricultural Statistics System - An Example

The following example of a process to build up an Agricultural Statistics System is relevant for the situation in Sweden and cannot as a whole be used in every other country. It is based on the building up of the new Swedish Statistical Farm Register, which is based on unique organisation identification numbers and personal identification numbers.
During the last decade, a system of administrative registers, named IACS, has been built up in order to run CAP. In order to reduce costs and response burdens when producing agricultural statistics, the information in these registers will be used. When creating a new Statistical Farm Register with good coverage, both these registers and other registers (for instance the "old" Farm Register) are used. To ensure the quality of the new established register questionnaires were sent out in August 2000 to all holdings in this new Statistical Farm Register. The aims were to check the population (holders, holdings) and at the same time collect additional data about animals.
The Swedish Agricultural Statistics System consists of the following four components.

6.1 A system of agricultural registers.

- IACS, which consists of the different registers for paying different kinds of premiums and subventions to the farmers, i.e. area aid for arable crops, livestock aid, environmental aid and slaughter premiums. Also a register for running milk quotas is included here. The different registers contain information about *farmers or farm enterprises applying for aid*, i.e. identification numbers, location, variables that determine the amount of aid or premium and also actual payments.
- Central cattle register (CCR) or data base for bovine animals, which contains information about *all cattle and their owners*, i.e. identification numbers of cattle and owners, location, birthday of the animals. As this data base in Sweden is used for paying special beef premiums, slaughter premiums and extensification premiums, also payments are registered in the data base.
- Registers for specialized farms that do not apply for any aid in CAP, such as specialized horticultural, pig producing or poultry farms.

6.2 Statistics from the agricultural registers.

After some register-statistics methodological work, *aggregated statistics* on crop areas, number of cattle, amounts of paid premiums etc. can be produced, which are based on IACS and CCR. It may be mentioned here that in the register for area aid for arable crops not only crops for aid but all crop areas of the farmers/farms are registered. It has been shown that in principle all arable land for the main crops in Sweden is included in this register. This is a kind of census statistics, as all crop areas, cattle and paid amounts are collected in the agricultural registers. In general the quality of the statistical variables is expected to be good, as most of the variables are used for paying the aid and there is a national control programme concerning these variables.
For this kind of aggregated statistics, the information in the registers need not be of high quality in every respect as it is not necessary to link the objects with other objects at micro-data level. The identification variables are thus not important and can therefore in this case be of lower quality. However, no statistics about farms can be produced, as farms are not the objects of the agricultural registers.

6.3 The new Statistical Farm Register.

When constructing the new Statistical Farm Register, the agricultural registers are linked together by linkage variables (usually the unique organisation or personal identification numbers) and ”register farms” are constructed on the basis of information about farmers or farm enterprises in IACS, owners in CCR etc. Here the register methodological work is very important when defining the new object "farm", which is usually not the object in the agricultural registers. This work also may include surveys among farmers/owners who cannot be identified as belonging to a certain farm, or farmers/owners for whom is not possible to decide whether they are working together on the same farm or not. It is also very important to follow up what has happened from one year to another to obtain complete coverage of the population in the statistical Farm Register.

Official statistics shall be of the proper quality, which is to say of sufficiently high quality in respect of the user's needs and of the costs involved. A basic requirement, however, is that good statistical practise shall be maintained, which means that the statistics are to be produced by means of recognized scientific methods, which meet requirements on quality, objectivity and reliability.

An important quality aspect concerns the possibility of obtaining reliable time series. For the time series to be meaningful there has to be comparability over time, i.e., continuity and robustness in the face of external changes. The statistics must also be adjusted in line with changes in the world around, otherwise they will not be relevant for the user. These two requirements – comparability over time and relevance at a particular time – cannot always be fulfilled at the same time.

Agricultural policy changes from time to time. The register system for agricultural statistics needs to be "self-sustaining" and independent, with use of IACS data but with control over the non-IACS variables and the non-IACS population.

Another important feature is to decide which time (day, month or year) the information is to bear reference to. For structure variables the information should bear reference to some certain day(s) while the information on payments on aid and premiums should bear reference to a whole year. Usually the information in the subsidy registers is up to date. On the other hand the information in the registers for specialized farms may be too old, why it is important to collect new information for the objects in these registers.

This Statistical Farm Register can be used

- for producing statistics,
- as a sampling frame for sample surveys or
- to complement survey data with register information.

6.4 Statistics using the information in the Statistical Farm Register.

a. Structure statistics and aid statistics.

The Statistical Farm Register can *directly be used for producing statistics* as structure statistics for crops and cattle. It is also possible to produce statistics showing the distribution of aid and premiums on different kinds of farms. This is a kind of census statistics, as all arable areas, cattle and paid amounts are collected in the agricultural registers and are included in the statistical register.

b. Additional structure statistics.

Using the Statistical Farm Register as *a sample frame*, a sample survey can be carried out to collect *additional information* in order to produce structure statistics concerning those variables that are not included or up to date in this register. Examples on additional information are number of pigs or number of poultry on the farms. This additional structure statistics will mainly be based on the sample survey, but it may also be possible to use the register information as complementary information in the estimation process.

c. Crop yield statistics.

For producing crop yield statistics, the Statistical Farm Register may be used as *a sampling frame for a regular sample survey* in collecting harvest data from the sampled farms. When estimating total yields for regions, the crop area information in the Farm Statistics Register can be used.

d. Book-keeping statistics (FADN).
For FADN, the Statistical Farm Register can be used as *a sampling frame* to choose the farms that shall be *included in the special FADN-register*. Structure information and information about aid and premiums will be transferred to this later register, which will be separated from the Statistical Farm Register.

e. Incomes of the Agricultural Household Sector (IAHS).
Household incomes statistics based on micro-level information can be produced after using the linkage variable personal identification number for farmers in the Statistical Farm Register in order *to define farm households* in the integrated incomes and tax register. In this later register the object set is all households in Sweden, for which the unique personal identification numbers of all household members are registered. The statistics on household incomes can be produced for households on different kinds of farms.

7. General administrative registers used for producing agricultural statistics

a. Education Register
The question on the manager's education was not asked in the agricultural census in Sweden 1999. Instead there will be an effort to produce statistics on that by using the data in the Education register (using the personal identification number for the manager as linkage).

b. Vehicle Register
In the agricultural census 1999 there was a partial non-response concerning machines. By using the information in the register on vehicles (using the personal code for the holder as linkage) the main part of the missing data could be found.

References

Selander, R., Svensson, J., Wallgren, A., Wallgren, B. (1998) *Administrative Registers in an Efficient Statistical System - New Possibilities for Agricultural Statistics? How Should We Use IACS Data?* Statistics Sweden and Eurostat.

Wallgren, A. and Wallgren, B. (1999) *Administrative Registers in an Efficient Statistical System – How can we use Multiple Administrative Sources?* Statistics Sweden and Eurostat.

Implications of Community Enlargement for European Agricultural Statistics

Derek Peare, Frans Kutsch Lojenga

Statistical Office of the European Communities (Eurostat)
Jean Monnet building, L-2920 Luxembourg
Tel.: +352-4301, extension 35188-D. Peare, 35912-F.J. Kutsch Lojenga
e-mail: charles.peare@cec.eu.int; frans.kutsch-lojenga@cec.eu.int

Abstract: The European Union (EU) is preparing for enlargement and is faced with negotiations with thirteen applicant countries. The role of statistics in these discussions is twofold; the incorporation of the candidates' statistical systems into the European Statistical System is an integral part of the negotiations while, at the same time, comparable statistics on economic and social aspects of the applicant countries are indispensable to the negotiation process. The purpose of this paper is to describe the processes by which these parallel aims are being tackled.

1. General introduction

1.1 Enlargement

At the turn of the new century the European Union is engaged on a major effort of enlargement and integration. Successive enlargements in the last three decades, the Single Market as well as Economic and Monetary Union have brought profound changes to the economic and political map of Europe. Both widening and deepening of the EU strongly affects the economies of Europe. This enlargement represents one of the greatest historical challenges yet faced by the EU.
Thirteen countries have applied for membership of the EU[1]. The number of potential candidates itself puts the EU to the test since never before have enlargement negotiations been conducted with such a large number of applicants. This enlargement has the potential to produce a population increase of just over 45%, which is some 170 million people (see table 1). Moreover, the political and economic background of these countries is basically different from the present 15 EU-Member States and from one another which implies a specific and divergent influence on their statistical systems.

1.2 Statistics

European statistics are the product of a close collaboration between Eurostat and the national administrations. Important inputs into the planning process come from the users of statistical information both within the Commission and other Community

[1] Bulgaria, Cyprus, The Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, The Slovak Republic, Slovenia and Turkey.

Institutions as well as from national administrations and the wider public both within and outside the European Union.

Table 1: *population and area in Candidate Countries*

	Population (millions)	**Area (in 1000 km²)**
Total EU-15:	**374,6**	**3.191,1**
Bulgaria	8,2	111,0
Cyprus	0,8	9,3
Czech Republic	10,3	78.9
Estonia	1,4	45,2
Hungary	10,1	93,0
Latvia	2,4	64,6
Lithuania	3,7	65,3
Malta	0,4	0,3
Poland	38,7	312,7
Romania	22,5	238,4
Slovakia	5,4	49,0
Slovenia	2,0	20,3
Turkey	64,3	769,6
Total 13 CCs:	**170,1**	**1.857,5**

On the basis of historical development, a logical division of tasks and the principle of subsidiarity, virtually all data collection and preliminary treatment is carried out in the Member States and results subsequently aggregated, analysed and disseminated by the Commission (Eurostat). The planning, maintenance and development of this complex set of relationships in an effective manner has led to the development of a set of interlinked networks. These administrative networks are supported by a range of technical tools which facilitate the sharing of tasks and rapid exchanges between partners.

The key challenges for statisticians in the enlargement process are the provision of data to support the negotiation process and the integration of the national statistical services into the European Statistical System.

1.3 Enlargement and statistics

Preparation for the enlargement involves intensive accession negotiations on a whole series of so-called chapters. The statistical system (both with Eurostat and with the candidate countries) itself is a chapter for negotiation. The proposals for EU-reform in Agenda 2000 require a common approach for comprehensive statistical data from the European Statistical System and the candidate countries. Clearly, Eurostat's role in co-ordinating the work on the statistical chapter and the compliance with all the (statistical) "acquis communautaire"[2] has become increasingly important.

[2] The 'acquis communautaire' comprises the entire body of legislation of the European Communities which has accumulated, and been revised, over the last 40 years. It includes

- the founding Treaty of Rome as revised by the Maastricht and Amsterdam Treaties
- the Regulations and Directives passed by the Council of Ministers, most of which concern the single market

2. Experiences with former enlargements

The successive enlargements of the EU have seen the founding Member States[3] joined by the accession of Denmark, Ireland and the United Kingdom in 1973, of Greece in 1981, of Portugal and Spain in 1986 and, most recently, of Austria, Finland and Sweden in 1995. As with the original Member States, the process of convergence of statistical systems has taken time to adapt the different traditions of national systems to the requirements of the Community system (which have themselves seen many changes over the years). On each occasion close collaboration between the services of the applicant countries, the Commission and those of existing Member States have served to ease the process of integration. The lessons learned from these experiences make a valuable contribution to the current process.

3. Need for data

3.1 Role of agricultural statistics

Agricultural statistics are generally regarded as very well developed in the context of EU statistics. They serve as an important information source for the (further) development and management of the common agricultural policy, which at present accounts for about 50% of the total EU-budget.
Agricultural statistics clearly play a major role in the framework of the enlargement negotiations with candidate countries. Enlargement will have a strong and diverse impact on the present EU Member States. It will strain the EU's economic resources, because the incomes of the new members are much lower than the EU-15 average, something that did not happen to the same extent in any of the previous enlargements. Also some EU policies, such as agricultural policy and structural policy, where the weight of the new members is high, may need to be adapted or reconsidered. The availability of comprehensive, reliable and timely information on agriculture is therefore crucial for providing, to the greatest extent possible, a solid basis for all decisions concerning the implementation of the common agricultural policy in this framework of the enlargement process. This is the case for market policy measures as well as for rural development policy. Solid agricultural statistics are also needed for other important purposes, like enlargement impact assessment, budgetary estimations, market prospects, etc. Relevant and timely data are required to assist the enlargement negotiations and to accompany the pre-accession strategy of the Commission.
Statistics is identified as a priority area in many Accession Partnerships and National Programmes for the Adoption of the Acquis Communautaire. Therefore, candidate countries require assistance in their efforts to comply with the statistical requirements of membership. The aim is then to ensure the multilateral transfer of know-how following a common methodology.

- the judgements of the European Court of Justice.

In practice additional, less formal, aspects are also treated as part of the acquis. See, for example 4.1 below.

[3] Belgium, France, Germany, Italy, Luxembourg and The Netherlands

To underline the importance of the assistance towards the candidate countries in statistics, some aspects may be particularly noted:

- Timing aspects: statistics require a long lead-time, continuous and steady work. For good quality data to be available by 2005 for reference years 2002-2004, appropriate efforts are required now;
- The acquis communautaire in statistics is constantly growing as a consequence of other Community and EU policies. Dynamic elements exist also in the current acquis; filling of gaps is still required in candidate countries;
- Co-ordination among statistical institutes and with other statistics producers: this is an important issue, as in each country a variety of institutions contribute to statistics production. Furthermore many statistics have international and bilateral aspects (external trade, balance of payments, ...);
- Strengthening of the statistical infrastructure ("Institution building");
- Economy of scale effects: a dominant reason, statistical co-operation has by its nature a multi-lateral character, resembling very much the operating mechanisms of Community Programmes. Delivery mechanisms in use, such as working groups, seminars, training etc. function on a multi-lateral basis. The participation of candidate countries increases the economy of scale effect;
- Several previous programmes have delivered good results, testified by external assessments. Also the maintenance of the achievements requires significant efforts.

3.2 Policy for statistics

Both the EU and the candidate countries need impartial, reliable and timely statistics on which to base policy decisions during the pre-accession phase and beyond. The need for further work to improve the comparability of their statistics with those of the EU was clearly expressed in the Commission opinions on the applications for membership, in Agenda 2000 and in subsequent progress reports.
The Governments of the Phare[4] candidate countries have signed Europe Agreements with the European Union and its Member States. Statistical data are needed to monitor these agreements. Co-operation in statistics is integral part of the Agreements.
Already in the pre-accession phase, e.g. because of participation in various Community programmes, a set of reliable high quality statistics is required, in steadily extending domains. National Statistical Institutes (NSIs) from all candidate countries have signed a Common Declaration of Statistical Co-operation with Eurostat. Through these agreements, NSIs have given their commitment to work closely with Eurostat and to make use of international and EU standards. They have declared their commitment to the objective of providing the same data that is currently provided by Member States with the same comparability and quality.

3.3 Programme

This Phare Statistical Co-operation Programme is based on a multi-annual indicative approach. It is proposed to finance statistical assistance and training activities in priority

---

[4] The Phare Programme (**P**oland and **H**ungary: **A**ction for the **R**estructuring of the **E**conomy) now covers Albania, Bosnia and Herzegovina, Bulgaria, the Czech Republic, Estonia, the former Yugoslav Republic of Macedonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia.

sectors, and specific sectoral projects (including projects of a pilot nature). It will aim to achieve the following specific results:

- Production of statistics in certain key areas comparable with EU standards for statistics, including the publication and dissemination of these statistics.
- Transfer of know-how concerning all EU requirements in statistics, and increase the coverage of Eurostat's regular databases. This will require close monitoring of the development of the EU statistical system, which is expanding rapidly, with (for example) up to 20 legal texts in preparation at any one time.

The activities to be undertaken to improve comparability and disseminate statistics in specific areas will be sectoral projects of a pilot nature. These projects will normally include an assessment of existing compliance with EU requirements, specific statistical assistance and training to assist countries develop appropriate techniques, and publication and dissemination of results. In all cases they are used to start new statistics or to reform an existing one aligning it to EU requirements. Where, and as far as, appropriate pilot surveys will be used to collect data. Projects will employ common methodology in all candidate countries and conform with EU standards. They will produce results in the form of useful statistics in priority sectors within a relatively short time period (months rather than years).

Pilot projects will be proposed by Eurostat, working closely with the other services of the Commission, according to the priority areas for the pre-accession period and deficiencies stated in the pre-accession process. Project proposals will be presented to the Policy Group for Statistical Co-operation in order to receive their commitment.

4. Current contacts and support in agricultural statistics

Against this background that agriculture is one of the most important sectors in most candidate countries (in terms of agricultural area, contribution to GDP, employment, etc.), the need to obtain a well-developed agricultural statistical system in which the candidate countries are fully integrated seems to be very clear.

4.1 Enlargement objectives for statistics

The EU enlargement objectives for agricultural statistics therefore are the following:

(a) Developing the appropriate statistical infrastructure, the so-called institution building. This implies reform of the official institutions and their reinforcement with new administrative and technical capacity, aimed at developing work structures, human resources and management skills.

(b) Investment in the "acquis communautaire", which is rather extensive regarding agricultural statistics. It consists of three elements:

- Community legislation (about 60% of the acquis communautaire), and

- "Gentlemen's agreements"[5] for important domains like agromonetary statistics, supply balance sheets, and land use (the other 40%)
- "Technical" documentation in the form of handbooks on methodologies and reports from Member States, workshops, etc.

The EU agricultural statistics might be regarded as a moving target of which the details of the present requirements are continuously changing. However, it is still useful that candidate countries develop their agricultural statistics in order to meet the present requirements (the "acquis"), because:
- the EU requirements form at least the starting point for tomorrow's needs;
- they form a comprehensive and harmonised entity which fits also into the world system;
- the data used in the accession negotiations will be based as much as possible on the European Statistical System.

4.2 Strategy in contacts and support

The key feature in the pre-accession strategy is the so-called ***Accession Partnership*** (based on Council Regulation no 622/98 of 16 March 1998) that seeks to mobilise all forms of technical and financial assistance to the candidate countries in a single framework. This framework covers in detail the priorities to be observed. As it is obvious that good statistics are essential for monitoring the CAP and also other Community policies, Eurostat has proposed to take statistics into account in these Accession Partnerships.
The Accession Partnerships are complemented by the ***National Programmes for the Adoption of the Acquis Communautaire*** ((NPAA), which give details on each country's commitments with regard to meeting the requirements of the acquis communautaire. For statistics, the NPAA's are prepared by the statistical authorities of the candidate countries. They are accompanied by annual programmes and quarterly plans, being the base for all technical (and financial) assistance to be provided.
The channel for EU's financial and technical support to the candidate countries has been the Phare programme. The ***Phare programming*** of statistical Co-operation is sub-divided into (1) National Phare programmes, and (2) Multi-beneficiary (or Multi-country) programmes.
Whereas national programmes concentrate mainly on the improvement of the statistical infrastructure including the provision of state-of-the-art hardware, multi-beneficiary programmes focus on the provision of comparable, timely and relevant statistics in certain priority areas, as well as the transfer of know-how concerning the statistical requirements of the EU.
Phare national programmes on statistics are presented mainly on the initiative of the (statistical) administrations of the candidate countries. They form part of a larger annual Country programme. The main weakness of such national statistics programmes is that they have to compete with programmes presented by other authorities, which also claim a fair share of the total amount available for the Country programme. In the beginning NSIs of the candidate countries had difficulties to find the support within their

[5] The term "Gentleman's agreement" is used to describe agreements which are legally non-binding but which underpin the production of statistics at EU level across a range of topics. Such agreements increase the flexibility of the system but their usefulness depends critically on the willingness of all partners to respect the understandings reached.

governments for the approval of the statistics programmes, but this has improved considerably. Especially in the area of agricultural statistics the joint preparation and co-ordinated lobbying of both the NSIs and the Ministries of Agriculture (institution building !) seems to be crucial during the preparatory phase of such annual Country programmes.
A third element of the pre-accession strategy is the so-called ***screening exercise***, which is based on a Council Decision requiring the Commission to submit to it an annual progress report on the (state of the) adoption of the legal acquis communautaire, for each country that has applied for EU-membership. The first screening on agricultural statistics took place in 1998 under the chapter "statistics" and under the chapter "agriculture". Gentleman's agreements are not included in the screening exercise. There are of course differences in the level of adoption among the candidate countries. The outcome of the screening exercise serves as a basis for the accession negotiations and is the political basis for the decisions to provide technical assistance.

4.3 Eurostat's support on agricultural statistics

Three kinds of ***documentation on the acquis communautaire*** have been prepared by Eurostat:
(a) the Statistical Requirement Compendium, containing a short description of each statistical module, including the gentleman's agreements (so-called "blue book");
(b) a compilation of all legal acts on agricultural statistics (the so-called "pink book");
(c) methodological documentation, i.e. handbooks with an extensive description of the methodologies used for the different modules.
As far as possible Eurostat tries to get a ***regular flow of data*** of agricultural statistics from the candidate countries, starting with agricultural production and prices. But as soon as possible also on structure and agro-monetary statistics. The data have been included in the NewCronos database under a separate subdomain.
Under the Phare multi-beneficiary budget for statistical Co-operation, Eurostat works closely together with the candidate countries on a number of ***pilot projects***. The first agricultural project was on agro-monetary statistics under the 1998-budget, which started in the beginning 1999 and comprises 4 subprojects. The budget for 1999 (some 4.3 M EUR for these agricultural pilot projects) has also 4 subprojects, of which one on farm structure surveys, two on agricultural production and one on environment. For 2000 and 2001 again a (smaller) number of pilot projects have been prepared for, but no final decision has yet been taken.
A first special meeting of the Agricultural Statistics Committee (ASC) on enlargement and ***technical assistance*** was held in October 1998, with the participation of the EU-Member States and the candidate countries. It provided transparency, clarified the political framework of Co-operation, pointed out the statistical needs of the candidate countries and it developed ideas for future technical assistance programmes on the basis of bilateral partnerships (twinning). The needs were broken down with two objectives:
(a) short term objectives, in the form of the organisation of so-called special workshops for candidate countries per statistical domain, organised directly linked to the regular Working groups. Results will help to include step-by-step topic related work to be integrated in the daily statistical work in the candidate countries, in co-operation with the Member States, and to elaborate some statements for the next screening exercises.

(b) long term objectives, in the form of the earlier mentioned pilot projects, bilateral technical assistance programmes between EU-Member states and candidate countries, training of human resources (stagiaires) and the so-called institution building (better and closer co-operation between NSIs, Ministries of Agriculture and Agricultural Economic Research Institutes).

Finally, all technical assistance towards the candidate countries is co-ordinated through several ***organisational network activities***. This concerns numerous bodies, like Eurostat, EU-Member States, institutions in the candidate countries, and international organisations like FAO, UN-ECE and OECD. The overall consistency of the agricultural statistical work will stay in line with the EU data requirements, guaranteed by the co-ordination and steering of the ASC.

5. Conclusion

The negotiation position of all candidate countries aims at acceptance of complete compliance with the "acquis communautaire" at the time of accession, no derogations or transition periods, under the condition of sufficient resources. The Common position of the Member States for all candidate countries so far is satisfaction and agreement, recognition of the importance of (agricultural) statistics, but also concern about the available resources capacity and about the sustainability of the (agricultural) statistical system.

When we take into account that there is a double dynamic situation in the form of a still growing "acquis communautaire" and a continuously changing compliance status, plus a still increasing need for detailed agricultural statistical data, then we can draw the following conclusions:

- The permanent and close monitoring of the compliance status becomes even more important. This is also in the interest of the candidate countries, as it is the only way to build sustainable confidence in their statistical system.
- The double role of statistics – an important chapter on its own and providing necessary data for the negotiations – must be recognised. They are part of the acquis, but also have to provide support for other chapters, of which agriculture is one of the important ones. In conclusion: negotiate with statistics, not about statistics.
- Resources are permanently a dominant issue. When candidate countries (and also sometimes Member States) do not (fully) comply, then the reason is not bad will, but normally they do not have enough or the necessary resources or infrastructure at their disposal. Providing the necessary assistance, therefore, is a prominent task of the European Statistical System, in terms of both financial aid and expertise, until the moment of accession.
- In doing this, the strategy has to be double: promote the allocation of national (Phare) programmes to the candidate countries and sectors in need, but also maintain a multi-beneficiary programme as main integration instrument for all the candidate countries. The European Statistical System has been quite successful in both approaches until now.

Statistics is not, and certainly should not be, the political focus. Agricultural statistics have the obligation to provide concrete, reliable and useful results.

6 June 2001

Invited Paper Session

CLASSIFICATION ISSUES FOR AGRICULTURE STATISTICS SYSTEM

Chair: M. Trant

Agriculture Far in the North – Different Kinds of Activities and Classification Problems

R. Poukka
Statistics Finland, PL 3C, FIN-00022, Helsinki, Finland
e-mail: riitta.poukka@stat.fi

Abstract: The purpose of statistics is to describe society, its activities, structures and changes in them. Statistics are like mirrors from which we can see an image of ourselves. How good and realistic this image is, is largely dependent on the classifications. Ultimately, the quality of the classification is determined by the users' needs: the classification should be relevant to various user groups' different purposes. According to users the standard classification of activities should be developed more clearly towards market activities and product groups. The 3 and 4 digit levels of the present industrial classification standard should be more disaggregated and the key criterion should be production output. Additionally, the information should be in a more structured form than at present, which would make possible specific classifications that would serve the needs of various types of users more flexibly and in more detail.

Keywords: agriculture, classifications, production line

1. Foreword

The purpose of statistics is to describe society, its activities, structures and changes in them. Statistics are like mirrors from which we can see an image of ourselves. How good and realistic this image is, largely depends on the classifications. One of the greatest challenges is international comparability. In this sense, industrial classifications are an excellent example of the questions involved in the application of international standard classifications and national classifications. This paper intends to examine how Finland's agriculture and forestry and its special features are described or not described in the standard classifications and what could be done to improve the classifications.

2. Special features of Finland's agriculture

Finland is the northernmost agricultural country. Its area is the sixth largest in Europe, it is very sparsely populated and the most forested country in Europe; 70 per cent of the land area is covered by forests. It is situated between the 60th and 70th latitudes. The distance from south to north is 1,100 km and due to Finland's northern location, natural conditions vary greatly from south to north.
The thermal growing season (the period with an average daily temperature in excess of +5° C) ranges from six months in the south to two to three months in the north. As the growing season is considerably shorter than in Central and Southern Europe, the types of cultivated crops must be suitable for our circumstances.

In Finland agriculture dominated until the 1950s. The proportion of the agricultural population was 41% in 1950, 17% in 1970 and just 7% now. New land was cleared fast for cultivation in place of the agricultural land ceded, i.e. lost in the war; at the same time a large number of new smallholders came into existence. The EU membership from 1995 onwards and adaptation to it brought about structural changes in agriculture: the number of farms has fallen considerably and the size of farms has grown.

Farms in Finland are mainly family farms. The size of arable land is typically under 20 ha. Natural conditions are reflected in production lines in that crop growing is focused on Southern Finland, pigs and poultry are farmed in Southern and Western Finland, while cattle farming is practised in Central, Eastern and Northern Finland. Specific features are fur farming and reindeer herding.

Finnish farms usually include forest as well. Only 5% of farms do not have forests at all. The majority of forests are thus owned by farmers and their descendants. Forest ownership has been and still is a significant asset item for farmers; income from sale of timber has been used to fund agricultural investments and to acquire a fairly large stock of agricultural machinery, in proportion to the arable area. Forestry continues to be the farmer's 'second job'.

3. Basic solutions of agricultural statistics production in Finland

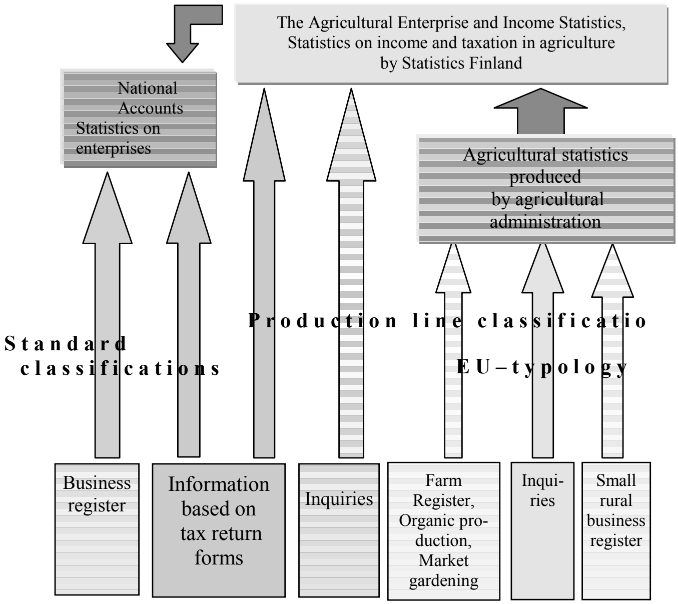
Two thirds of Finland's national statistics production is conducted by the national statistical agency, Statistics Finland, and the remaining one third outside it. With regard to agricultural statistics, Statistics Finland's task is to compile statistics on agricultural income and taxation and agricultural indices, while the Ministry of Agriculture and Forestry and the administrative organisations working under it produce other statistics on agriculture, forestry, fishing and so on.

Statistics production in Finland is based to a great extent on statistical or administrative registers. The Statistics Finland Act (24.1.1992/48) gives statistical authorities the right to obtain data needed for statistics compilation from administrative registers. A great deal of statistics are thus entirely based on registers. – Several registers are also source materials for statistics describing agriculture. They are used to define the units belonging to the target population and their central characteristics. The key registers forming the basis of statistics productions are as follows:

- The Farm Register, now the Rural Business Register maintained by the Information Centre of the Ministry of Agriculture and Forestry. The register covers all holdings with at least one hectare of arable and horticultural land in management. Besides the Rural Business Register, registers covering certain sub-sectors include registers on horticultural enterprises, organic production and small agricultural enterprises.
- Information on the taxable income and expenditure, and assets and liabilities of units engaged in agriculture is obtained from the Tax Administration registers. The register comprises all recipients of agricultural income, regardless of whether it concerns their principal activity or secondary income.
- The Business Register maintained by Statistics Finland is a statistical register and covers all enterprises, entrepreneurs and the like, excluding farms. Farms are included for the time being in the register only in case the farm is also engaged in small-scale business activities. Therefore the Business Register is not very significant to the compilation of agricultural statistics.

As is apparent from the above, the uses and target populations of the registers concerned are different. The following figure 1 illustrates further the division of labour in statistics production and the acquisition of basic data roughly.

Figure 1: *Division of labour in statistics production and the acquisition of basic data in Finland*



In Statistics Finland's Business Register an industry code is given for enterprises and other units engaged in economic activities in accordance with the national Standard Industrial Classification (SIC95) based on the standard classification of the European Communities, NACE Rev.1.

The statistics of the agricultural administration, the agricultural enterprise and income statistics and the statistics on agricultural income and taxation prepared by Statistics Finland make use of the agricultural administration's national classification of production lines together with the Community typology for agricultural holdings.

3.1 Principal differences of classifications

The Industrial Classifications – ISIC, NACE and SIC95 – are based on the framework of the national accounts. The classifications cover all economic activities, the whole agricultural sector as well. In contrast, the agricultural administration's classification of product lines and the Community typology are constructed to describe only farms and production activities on farms. That's why the primary administrative classification – the national classification of product lines – does not recognise mixed farming, i.e. growing of crops combined with farming of animals, or reindeer herding and fur farming. The latter ones are also left out of the Community typology while the typology includes mixed crops and livestock and their many variations.

Product classifications are also different. Statistics Finland uses the EU's product classification CPA, which follows to a large extent the UN's product classification, CPC. Compared to them, the product classification of the agricultural administration is very detailed. For example, it classifies the grain varieties separately and further splits it into autumn and spring grain; the potato has nine groups: food potato, food industry potato, starch potato, etc. Animals are similarly grouped by species and further by age and/or weight and other such criteria. As the administration's detailed classification can always be used to form groups according to CPA, product classifications are not necessarily a problem and they will not be examined in more detail in this paper.

Target populations

In the agricultural enterprise and income statistics and in the agricultural income and taxation statistics produced by Statistics Finland the basic target population is an agricultural holding engaged in agriculture and forestry with at least one hectare of arable and horticultural land under cultivation. A further condition is that agriculture is taxed according to the Agricultural Income Tax Act; the taxation criterion excludes cases where farms are engaged in such activities related to agriculture and forestry that have to be considered as a separate business activity. The product lines of farms are defined on the basis of income received from the products.

In the statistics compiled by the agricultural administration the target population includes all active farms with at least one hectare, all horticultural farms and such forest estates that also have fields under cultivation. The production line is obtained from the register and it is based on the notification made by the farmer in connection with the Agricultural Census. By definition, the Farm Register covers only farms and thus such forest estates that have no cultivated fields are excluded from the register. Therefore, the essential part of Finnish agriculture, forestry, is left out of the both administrative classifications.

3.2 Organic production, peat, small agricultural enterprises

There are several special questions related to the industrial and product classifications – definitions, demarcation problems, etc. – but we will discuss only a few of them here.

Organic production

Organic production has been practised only a relatively short time in Finland. It expanded with the EU membership, which made it possible to decrease the profitability difference. At the moment, about 6% of arable land is organically produced, which means that Finland is one of the top organic producers, in relative terms. In the Rural Business Register organic production is still one of the undivided production lines but it will be divided in the near future into crop growing and animal farming. As we know, industrial classification standards do not class organic production into its own group.

Peat

Finland has the highest number of marshes in Europe. One third of the area is marshland. We are the largest peat producers in the EU. We have long been stripping peat from our marshes and drying it for fuel, briquettes and pellets, horticultural peat being the other mode of use. Peat processing is mainly a business activity, but peat has also been processed to a certain extent as a small-scale business or for domestic use at farms, since marshes similarly as forests are primarily owned by private farmers.

The processes of peat stripping and agglomeration differ crucially from mining and quarrying. Peat marshes are also renewable, although slowly, but they can also be restored to other production use. For these reasons we would rather see peat extraction and agglomeration as part of agriculture and forestry than mining and quarrying. An initiative of such classification revision has been made to Eurostat for the updating of the NACE 2002; it will not be realised in this context, however. The decision made by the EU in November 2000 on including peat as a slowly renewable energy source can be held as supporting this view. According to the decision, peat could be classified as biofuel if its extraction from marshes does not exceed its annual growth.

Small enterprise activities on farms

In Finland a register on small rural enterprises has been established for research purposes kept by the Agricultural Economics Research Institute. Its purpose is to monitor and examine small rural enterprises, one part of which are businesses operating in connection with farms. Small enterprises are defined to be such that employ 0.5 to 20 person-years, are privately owned, have one establishment and whose turnover is in excess of the set threshold value.

This means that the activity of farms is now to be surveyed in more ways than it was possible earlier. Farms are more often than before engaged in other activities, which supplement farming and provide additional income: farm produce sale, farm accommodation, renting of machinery, snow removal from summer cottages, and so on.

The matter has been under consideration at Eurostat as well. From the viewpoint of industrial classification, such small-scale enterprises directly connected to agriculture, e.g. to the processing of the farm’s own production and related sales, has no influence on the definition of the primary activity, that is, it is classified as agriculture.

4. Do standard classifications describe our agriculture adequately?

The validity of the classification is the sum of several factors. The classification has to be exhaustive, the criteria used must be unambiguous and mutually exclusive and the theory or the frame of reference behind it has to be so effectual that the classification as a tool operates adequately under different circumstances and endures well the changes of time. Ultimately, the good quality of the classification is determined by the users’ needs: the classification should be relevant to various user groups’ different purposes.

The classification is a tool for arranging the data. The information from other than quantitative classifications is usually compressed into a single code. We all know that no industrial classification can describe the whole spectrum of the activities of an enterprise or a farm by one code.

Validity of industrial classifications

Standard classifications – ISIC, NACE and SIC95 – in principle cover all the activities related to agriculture, unlike the Finnish national classification of production lines or the European Community typology that are restricted to crop growing and animal farming. However, industrial standards are not very apt to describe Finnish agriculture in micro level statistics. There are mainly two problems.

First, the more exotic production lines of the Finnish agriculture – fur farming and reindeer herding – have to be classified in ISIC and NACE under residual group ”Other animal farming”. They are brought to the fore in the national classification, SIC95. Another shortcoming is that the industrial classifications do not describe typical Finnish agriculture so that the classifications would take it into account that the majority of farms also practise forestry. That is, "combined crop growing/animal farming and forestry" could form a separate group in the industrial classification similarly as it has a group for growing of crops combined with farming of animals.

The same is repeated in the EU's product classification standard, CPA. The picture of Finnish agriculture is not actually made more specific through it, although CPA has to be seen as a specification on the industrial classification, NACE. CPA does not recognise reindeer meat and furskins are not separated according to animal. As a small detail it could be here mentioned that CPA does not either include the main products of Finnish lake fishing, freshwater fishes (northern pike, perch, pike-perch, etc.) that are also fished in our seas that are low salt brackish waters.

In rough summary we can say that with respect to agriculture and forestry (main groups A and B), industrial standards are better applicable to macro level statistics, such as national accounts than to micro level agricultural statistics. In international comparisons the image of Finland and its special features are not much taken into consideration in standard classifications.

User needs

If one asks users or compilers of agricultural statistics why the standard industrial classification is not used in statistics describing agriculture, the clear answer is that the standard classification of official statistics is too coarse and agricultural production is not described from the viewpoint of markets. The information needs of those using agricultural statistics are based and centred on agricultural production (production structure, volume, prices, etc.) and on market activities by product and product group.

The statistics production of the agricultural administration is naturally harnessed to produce the necessary data for regional and structural policies aiming not only to improve agriculture and its operational conditions but also to develop more widely rural businesses and to prevent depopulation of the countryside. Key tasks are preparation for decisions related to agricultural policy and business subsidies and production of the information needed as well as assessment of food markets, self-sufficiency and so on.

Both data production and data needs are centred on production: structure and volume of production and price data. The concept of principal activity used in industrial classifications (NACE and SIC) is not suited for such examination. This is the main reason why the industrial classification in agricultural statistics is displaced by the classification of production lines and the Community typology.

5. How classifications could be improved in the future?

The differences between the industrial classification and the production line classification have been seen problematic. This is not just Finland's problem but a more general dilemma for the whole EU.

To improve consistency with production line classifications

The standard classification of official statistics should be developed more clearly towards market-oriented production line and product group classifications. In Europe it is important to improve consistency with the EU typology as the data on the Member States have to be produced anyway following that classification. Finland plans to use in future the EU typology alongside the national main production line classification. This causes no particular problems because both of the classifications have basically the same orientation. In this way, the administrative classifications could be brought closer to the industrial classification, especially as concerns the classification of combined production lines, although this would probably not be enough.

Disaggregation

For the development of the industrial classification, the shortcomings reported by the users mean that the 3 and 4 digit levels of the present industrial classification standard should be disaggregated and the key criterion should be production output. It is easy to add breakdowns to the national industrial classification in connection with updating, in which case the classification would specify the picture given by statistics on the national level. But this is not a sufficient solution for specifying the picture internationally. However, we may have to accept the fact that it may be hard to find a place for the special features of a small country in pan-European or even global industrial classifications. For this reason it should be examined what we could achieve by specifying product classifications.

The natural way to progress would be to make CPA more specific, as it is a classification binding the EU Member States and it is more detailed than CPC. It would be rational to specify CPA also because CPA acts directly as a specification for the industrial classification; CPA is a structured classification where each commodity is linked to only one industry group. However, this structured linking (binding commodities to industry groups) has been criticised as well. It has been hoped that the products resulting from different processes could be linked through each process to the industry behind it, which means that a CPA commodity could be combined to more than one industry group. However, Eurostat's NACE/CPA working group wants to stick to the principle that one product is combined to one industry group only.

Building block information

It is evident that we need to include more information in relevant business registers. The information should be in a well structured form, which would make possible specific classifications that would serve the needs of various types of users more flexibly and in more detail. – The ongoing Clamour project (**CLA**ssifications **MO**delling and **U**tilities **R**esearch) examines the prerequisites for the development of more multidimensional classifications applicable to more uses and for structuring and modelling the basic data needed for forming classifications. The Clamour project is one of the projects related to the EU's Fifth Framework Programme for Research and Technological Development concerning the promotion of the information society. The project aims to produce such information on the fundamentals of classifications and user needs that could be utilised in the industrial classification revision for 2007, for example.

The objective of the building block method is to develop a model to describe the activities of a business. This part of the Clamour project is handled by the Dutch, who define their task as follows: "First a model is presented that is rich enough to fulfil most anticipated users needs. Then this model is simplified in such a way that the information needed can be obtained from businesses with a realistic effort. An obvious application of the model is its use as an activity description in a multi-purpose single business register. Another application is its use as a conceptual basis for a system of activity classifications". This means that the information collected on the activity of an enterprise or a group of enterprises is dispersed into natural parts as concerns the enterprise's activity regardless of its formal organisation. More information on the Clamour project is available at: www.statistics.gov.uk/methods_quality/clamour.asp.

At the same time as the users want to have more detailed and multi-purpose classifications, they also value common standards. The structure and criteria of the classification should be the same on the national level, in the whole continent and all around the world. This speaks strongly for the preservation of the harmonised system.

References:

The Agricultural Enterprise and Income Statistics 1998 (2000). Statistics Finland, Helsinki.

Farm Register 1998 (2000). Information Centre of the Ministry of Agriculture and Forestry, Helsinki.

CLAMOUR (CLAssifications MOdelling and Utilities Research). EU's Fifth Framework Programme for Research and Technological Development concerning the promotion of the information society. *See:* www.statistics.gov.uk/methods_quality/clamour.asp

CPA (1993) = Statistical Classification of Products by Activity in the European Community, *Council Regulation (EEC) No. 3696/93 of 29 October 1993 on the statistical classification of products by activity (CPA) in the European Economic Community*, OJ L 342 of December 1993.

Eurojyvä, Maatilat Suomessa ja EU:ssa (2000), Information Centre of the Ministry of Agriculture and Forestry, Helsinki. (In Finnish)

EU typology, *Commission decision (EEC) No. 377/85 of 7 June 1985 on establishing a Community typology for agricultural holdings (amendments 376/94/EC, 393/96/EC and 725/99)*.

ISIC = International Standard Classification of All Economic Activities (1990), Third Revision, United Nations, Statistical Papers, Series M, No. 4, Rev.3, New York.

Bijwaard H., van der Hoeven, J., van Hooff, H., Kroese, B., Willeboordse A. (2000): A model to describe the activities of a business, Statistics Netherlands, Voorburg.

NACE = Nomenclature générale des activités économiques dans les Communautés Européennes (Rev.1) (1993), Council Regulation (EEC) No. 3037/90 of 9 October 1990, as amended by Commission Regulation (EEC) No. 761/93 of 24 March, OJ L 83, 3 April 1993.

Rainer, Norbert: The Revised System of International Classifications. *Internal document,* Eurostat Data Shop Handbook, 1995.

Rantamäki-Lahtinen, Leena: Maaseudun pienyritysten rakennekehitys (2000). Agricultural Economics Research Institute, Finland. Research reports 245, Helsinki. (In Finnish)

SIC95 = Toimialaluokitus 1995 (1999), Toinen tarkistettu painos. Tilastokeskus, Helsinki. (Standard Industrial Classification 1995 (1999). Statistics Finland, Helsinki.

Statistics Finland Act, 24.1.1992/48

Suomen maatalous ja maaseutuelinkeinot 1999/2000 (2000). (Agriculture and rural activities), Publications 95, Agricultural Economics Research Institute, Finland, Helsinki. (In Finnish)

Annex. Comparison of the standard classifications (NACE and ISIC) and the Finnish national version (SIC95) and the national production line classification

ISIC Rev.3	NACE Rev.1	SIC95	Holdings in the Farm register by production line
01 Agriculture, hunting and related service activities	**01 Agriculture, hunting and related service activities**	**01 Agriculture, hunting and related service activities**	
011 Growing of crops; market gardening; horticulture	**011 Growing of crops; market gardening; horticulture**	**011 Growing of crops; market gardening; horticulture**	**Plant production**
0111 Growing of cereals and other crops n.e.c.	0111 Growing of cereals and other crops n.e.c.	0111 Growing of cereals and other crops n.e.c.	Crops production
0112 Growing of vegetables, horticultural specialties and nursery products	0112 Growing of vegetables, horticultural specialties and nursery products	0112 Growing of vegetables, horticultural specialties and nursery products	- Special plant cultivation - Horticulture in the open - Cultivation in greenhouses - Other plant production - Organic production
0113 Growing of fruit, nuts, beverage and spice crops	0113 Growing of fruit, nuts, beverage and spice crops	0113 Growing of fruit, nuts, beverage and spice crops	
012 Farming of animals	**012 Farming of animals**	**012 Farming of animals**	**012 Farming of animals**
0121 Farming of cattle, sheep, goats, horses, asses, mules and hinnies; dairy farming	0121 Farming of cattle, dairy farming	0121 Farming of cattle, dairy farming	- Milk production - Other livestock farming - Beef production - Other livestock
	0122 Farming of sheep, goats, horses, asses, mules and hinnies	0122 Farming of sheep, goats, horses, asses, mules and hinnies	
0122 Other animal farming; production of animal products n.e.c.	0123 Farming of swine	0123 Farming of swine	- Pig husbandry - Pork production
	0124 Farming of poultry	0124 Farming of poultry	- Egg production - Poultry meat production - Other poultry husbandry
	0125 Other farming of animals	0125 Other farming of animals	- Sheep husbandry - Goat husbandry - Horse husbandry - Other husbandry
		01251 Fur farming	
		01252 Reindeer farming	
		01259 Raising and breeding of other animals	

013 Growing of crops combined with farming of animals (mixed farming)	**013 Growing of crops combined with farming of animals (mixed farming)**	**013 Growing of crops combined with farming of animals (mixed farming)**	
014 Agricultural and animal husbandry service activities, except veterinary activities	**014 Agricultural and animal husbandry service activities, except veterinary activities**	**014 Agricultural and animal husbandry service activities, except veterinary activities**	
	0141 Agricultural service activities	0141 Agricultural service activities	
	0142 Animal husbandry service activities, except veterinary activities	0142 Animal husbandry service activities, except veterinary activities	
015 Hunting, trapping and game propagation including related service activities	**015 Hunting, trapping and game propagation including related service activities**	**015 Hunting, trapping and game propagation including related service activities**	
02 Forestry, logging and related service activities	**02 Forestry, logging and related service activities**	**02 Forestry, logging and related service activities**	
	020 Forestry, logging and related service activities	**020 Forestry, logging and related service activities**	
	0201 Forestry and logging	0201 Forestry and logging	- Forestry - Farms with field under cultivation - Farms without field under cultivation
		02011 Reafforestation	
		02012 Silvicultural and forest improvement work	
		02013 Timber harvesting	
		02019 Other forestry and logging activities	
	0202 Forestry and logging related service activities	0202 Forestry and logging related service activities	
05 Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing	**05 Fishing operation of fish hatcheries and fish farms; services activities incidental to fishing**	**05 Fishing operation of fish hatcheries and fish farms; services activities incidental to fishing**	
	050 Fishing operation of fish hatcheries and fish farms; services activities incidental to fishing	**050 Fishing operation of fish hatcheries and fish farms; services activities incidental to fishing**	
	0501 Fishing	0501 Fishing	- Professional marine fishery - Food fish production
	0502 Operation of fish hatcheries and fish farms	0502 Operation of fish hatcheries and fish farms	

Agriculture and Fishing Activities in the Pacific – The Special Classification Needs of Small Island Economies

Garth Parry
Secretariat of the Pacific Community, BP D5, Cedex 98848 Noumea
New Caledonia
Tel: (687) 260145 Fax: (687) 263818.
e-mail: GarthP@spc.int

Abstract: In fragile economies it is vital that planners have access to an appropriately detailed level of reliable information. This is true in particular for industries such as agriculture and fishing, where factors such as economic pressures and rapidly changing technologies have the potential to destroy delicate balances if changes are not managed well. However, the level of detail available is governed by the classifications used in the collection of the source data. A potential problem here is that international classifications are strongly influenced by highly developed economies that dominate global GDP and which tend to have very different characteristics from most Pacific Island economies. This paper aims to demonstrate the need for a regional classification of agricultural and fishing activities in the Pacific, and so question generally whether there should be more special purpose classifications.

Keywords: Fishing, Agriculture, Classifications, Pacific, Islands, PICT, Region

1. Introduction

Irrespective of their size or stage of development, individual countries and territories throughout the world possess some specific characteristics that differentiate them from other countries and territories. Further, when countries and territories are grouped into "regions" the same situation applies – each region possesses some characteristics which differ noticeably from those of other regions. This national and regional heterogeneity represents an enormous challenge for the statisticians developing international classifications: how to account as fully as possible for global diversity, but to do so in a manageable framework? The nature of this problem is such that there will always be instances where "international standards" fall short of the ideal for more narrowly-defined studies such as, in the context of this paper, detailed research relating to Pacific[1] island countries and territories (PICTs). But does this imply that international standards are somehow inadequate, or is it simply that we need to develop additional special purpose tools? This paper aims to support the latter interpretation.

It is not being suggested that the Pacific is any "more unique" than other regions, nor that international classifications should be changed to better reflect our needs. Rather, this paper aims to explore the special classification needs of the Pacific for agricultural

[1] The term "Pacific" is often used very loosely to describe a range of countries and territories in the Central and Western Pacific. In this paper it is used to refer to the 22 Pacific Island Countries and Territories (PICTs) who are members of the Pacific Community, their EEZs and the adjacent high seas areas.

and fishing activities, and thereby illustrate the need for "regional" classifications which can be used to supplement international classifications and so help to improve the consistency and comparability of work done throughout any particular region.

2. The use of classifications in Pacific statistics

In preparing for any study involving statistics, the general question should always be asked: are international standard classifications and concepts relevant for the purpose of the exercise? Clearly, there are immediate advantages for using existing classifications: international comparability, ready availability of well-defined concepts, and easy access to expert advice are some that spring readily to mind. But the advantages of using existing classifications need also to be weighed against the alternative benefits of using measures specific to the needs of each study. This issue can also be viewed from a regional perspective – is there sufficient commonality in studies being done throughout a region to justify the development of classifications specific to that region?

Let us consider the *International Standard Industrial Classification of All Economic Activities* (ISIC) as a test case. The latest version of ISIC was released by the UN in 1990, it has 292 component "classes" of economic activity, and its development by the UN Statistics Division involved extensive consultation with expert statisticians throughout the world. Nevertheless, it is inevitable that the sort of global decisions that underlie developments such as ISIC require at least some degree of compromise. An obvious concern for regions with a relatively small contribution to the global economy is that "international" standards may not accurately reflect their own circumstances.

Indeed, the need to sometimes adapt international classifications for other purposes is often recognised by the classifications themselves. For example, the notes to the 3rd revision of the ISIC include Chapter III: *Application of the classification,* with separate sections dealing with this issue: *B. Use of ISIC in establishing related national classifications*, and *C. Expansion or contraction of ISIC*. It is only a short step from developing national classifications to developing regional ones.

There are a number of factors which suggest that agricultural and fishing activities in the Pacific may need additional work on their systems of classification:

- agriculture and fishing are dominant sectors of many PICT economies, far more so than in many other regions
- there have been many diverse studies involving agriculture and fishing in the Pacific, and there is no indication that this will change
- agriculture and fishing seem to be relatively poorly served by existing classifications such as ISIC.

3. The importance of agriculture and fishing activities in the Pacific

The Pacific is a large region physically, and it has some quite striking characteristics. One of these is that agriculture and fishing are a far more dominant feature of most PICT economies than they are in larger, more developed economies. For example, Australia has a huge landmass and vast shoreline but *Agriculture, forestry and fishing* activities accounted for only 3% of its total GDP in 2000. The most recent comparable

PICT data is for Samoa, where *Agriculture, forestry and fishing* activities accounted for over 17% of GDP in 2000. Even though the Samoan economy has experienced several years of very strong growth in its services sector, the agriculture and fishing share of GDP is still six times as important in Samoa as it is in Australia. And that is measuring its relative importance only in monetary terms, which does not make any allowance for the very significant cultural and social importance attached to traditional food production and consumption.

There are many other striking examples of these differences – Kiribati has an EEZ roughly the size of mainland USA, but has a population of only 91,000 people living on 811 square kilometers and consuming on average over 180kg of tuna per annum. Fishing activities alone account for over 25% of Palau's total GDP. The value of the Western Pacific tuna catch is around US$2 billion annually, equivalent to more than 10% of the combined GDP of all the countries in the region. At the height of its production, the tuna canning industry (manufacturing, but fishing-related) in American Samoa supplied 70% of the USA market for canned tuna and employed 5,000 people, around 8% of American Samoa's entire population. Agriculture and fishing in combination account for over 20% of GDP in many PICTs. These are just a few examples highlighting the importance of fishing and agriculture to the region, as well as it's somewhat unusual geography and structure.

Given the sorts of relativities outlined above it is clear that the current classification by ISIC of *Agriculture, forestry and fishing* activities does not reflect the economic structure of a typical PICT. One simple example: in ISIC all fishing activities are classified to a single class (*0500, Fishing*), representing just 0.3% of the 292 classes allocated. (The previous version of ISIC did have two classes for *Fishing and related activities,* but these were collapsed into a single class with the introduction of Revision 3.) Agriculture is slightly better served with four individual classes, but this is still clearly insufficient for PICT-oriented research and analysis. The fact that ISIC has evolved into its present form implies that larger, developed economies do not need more detailed industry classifications for fishing and agriculture. This situation is unlikely to change, so it would be unwise for the PICTs to sit back and hope that ISIC will evolve of its own accord into something better suited to their needs.

4. Classification of agriculture and fishing activities in the Pacific

Most agricultural censuses and surveys in the Pacific have been essentially commodity-based, with little or no attempt to classify producing units to a predominant activity. But there are often very large differences in the nature of the activities underlying the production of the different crops. If one were to classify Pacific agricultural producers according to ISIC, analysts would quickly run into the problem that activities which are regarded locally as quite diverse would be classified within the same class. The same is true for fishing, although here at least many surveys have focused on particular types of producers, generally developing their own sub-classifications in the process. And, regrettably often, the various estimates of agriculture and fishing production have been combined into one aggregate for publication, something which has consistently frustrated analysts in the region who try to obtain data on one or both of these industries from publicly available databases.

Having said that ISIC does not fully serve the Pacific's needs for agricultural and fishing classifications, it is only fair to then put forward views on what those needs are. The following sections attempt to provide draft classifications that can be used as the basis for discussion throughout the region, leading hopefully to some form of a regional standard.

4.1 Classification of agricultural production in the Pacific

There are a number of major crops which are common throughout the Pacific region, and these can be used a starting point for suggesting which agricultural activities are deserving of separate classification.

Given the prevalence of coconuts throughout the Pacific, ISIC 0113: "*Growing of fruit, nuts, beverage and spice crops*" provides a good example of the need for further disaggregation. A family/producer with most of its agricultural effort aimed at a coconut plantation would be classified to the same ISIC class as would another producer whose main crop was citrus fruit, or pawpaws, or breadfruit, or avocados, all of which are important crops in certain PICTs. But the material inputs and growing techniques of these latter crops are different from each other, as well as being significantly different from those of a coconut plantation, and so should be recorded separately in any exercise aimed at measuring structural relationships within the agriculture sector. If one were to aim for an industry classification for agricultural activity in a typical PICT it would seem sensible to break ISIC 0113 into a number of finer-level activities. (There would still be the ever-present problem of classifying "mixed farming" units in such cases, but that is a separate issue.)

Another crop which is very important throughout the Pacific is taro, a high starch root crop. This is a very highly-valued food and is a dominant part of the staple diet of many Pacific Islanders. In 1994 Samoa was struck by the taro leaf blight and there was major national consternation – there was a severe disruption to national food supplies, with consumers having to adapt to the much less preferred ta'amu, supplemented by significantly increased imports of rice. (There was even a rumour circulating for a while that the blight had been introduced deliberately by rice importers – fortunately this was subsequently disproved.) The blight also led to major disturbances in the Samoan CPI, as taro by itself had a weight of over 5% and the price of the taro that was still available increased manifold as the supply dwindled to nothing.

There is a story involving taro that is of some relevance here. When a very prominent Pacific Islander was young he consulted a British doctor about his diet. On hearing that taro was a prominent part of the patient's diet the doctor asked what "taro" was. When the patient replied that it was a vegetable the doctor encouraged him to eat as much of it as he liked as it would obviously be good for him. But given the very high starch content of taro, some vigorous adherence to the doctor's well-intended advice led to the patient gaining a great deal of weight very rapidly. The story has been told widely throughout the region, mainly as a humorous reflection on the lack of understanding of the Pacific, and it does illustrate two points of relevance to the classification debate. The first and more obvious one is that the knowledge and interpretations of international experts are not always appropriate for the Pacific. The second may reassure some of the classification experts: from a Pacific perspective ISIC appears to be entirely justified in its recommendation that "Roots and tubers with a high starch content" be separated out from *Class 0112: Vegetables*.

ISIC 0111, "*Growing of cereals and other crops n.e.c.*" provides another good example of the need to disaggregate for Pacific conditions. This class includes activities which are quite common throughout the Pacific but are as diverse as the growing of peanuts, tobacco, yams and sugar cane. In well-developed economies the growing of these crops is not sufficiently significant to justify the creation of separate industry classes, but in various PICTs one or more of these activities definitely deserve treatment in their own right. For example, an agricultural survey in Fiji would almost certainly want to differentiate between the operations of sugar cane producers and those of major taro growers. The two crops are grown and harvested in very different ways, and both are sufficiently major export earners to be of interest in their own right. But ISIC 0111 includes both of these activities in 0111 as "Roots and tubers with a high starch content" have been specifically excluded from *0112: Vegetables*.
As an initial suggestion for an appropriate level of disaggregation, the growing of the following crops is common to a range of PICTs, the techniques involved are extremely heterogeneous between groupings, and each is sufficiently important in at least one PICT to be worthy of separate classification. The draft classification aims for some hierarchical link to the existing ISIC.

01111	Growing of potatoes, yam, sweet potatoes, taro, and similar crops
01112	Growing of cereal crops (other than in 01111)
01121	Melons, squash, etc (of particular relevance for Tonga which supplies the Japanese market, most being used for tempura batter)
01122	Other vegetables eg, tomatoes, capsicums, cabbage
01131	Coconuts
01132	Citrus fruits
01133	Other fruits eg, pawpaw, breadfruit
01134	Sugar cane
01135	Beverage crops eg, coffee, cocoa beans
01136	Other crops, including spice crops eg, vanilla, ginger (and both of these have considerable importance in individual countries)

Looking at the farming of animals, the relevant ISIC class is 0122: "*Other animal farming, production of animal products n.e.c.*" An immediate limitation for Pacific farming is that this single class includes both pigs and poultry. Most PICTs have some organised poultry/egg production and ***every*** PICT has extensive pig farming, although much of it is relatively informal. Cattle farming is a major activity in only a few PICTs, but there are a number of countries trying to develop it and who would have a keen interest in separately identifying it in any agricultural survey. The differences between these activities suggest a need for separate classes of economic activity. Possibly:

01211	Farming of cattle, sheep, etc
01212	Pig farming
01213	Poultry farming, including egg production
01214	Other animal farming; animal products n.e.c. (this could include the crocodile farming being developed in Palau)

The existing ISIC classes 0130 (mixed farming), 0140 (agricultural services) and 0150 (hunting and trapping) could probably be usefully retained in their existing formats, with their codes expanded for consistency to 01300, 01400 and 01500.
In total this would provide 17 categories of agricultural production, a significant increase from the four classes in the current ISIC, but a far more realistic basis for classifying diverse activities which provide as much as 20% of GDP in some PICTs.

4.2 Classification of fishing production in the Pacific

Looking at fishing, it is clear that the single ISIC class 0500 "*Fishing; operation of fish hatcheries and fish farms; service activities incidental to fishing*" amalgamates a range of activities which are each of considerable interest to the region. As noted previously, Revision 2 of ISIC did separate out the two broad categories, "*Ocean and coastal fishing*" and "*Fishing in inland waters, fish hatcheries, cultivated beds, fishery service activities*", but these were merged into a single class in Revision 3. This suggests strongly that any future revisions of ISIC are unlikely to break 0500 into more than one class – as such there is a strong case for developing a "regional standard" set of fishing codes.
Some of the statistics quoted previously suggest a need for separate classification of tuna fishing – but do the other activities covered by ISIC 0500 need separate identification? The answer is most certainly that they do. Not only are they extremely diverse in nature but there is strong (and growing) interest in monitoring activity in many aspects of fishing, including some which may be relatively minor at present but are hoped to develop in the future. For example, while Kiribati is very strongly influenced by tuna fishing – local per capita consumption of tuna is estimated at over 180kg per annum, and the value of the foreign-owned longline catch exceeds the total GDP – there is also very keen interest in the tourism implications of bonefishing on Kiritimati Island. If one were developing a fishing classification for Kiribati the obvious starting point would be separate treatment of long-line fishing, artisinal fishing, and recreational fishing.
The range of activities in 0500 includes many which are of great importance to particular PICT economies:

- *Pearl farming* is a dominant feature of economic activity in the Cook Islands and French Polynesia
- New Caledonia has a major aquaculture project for *prawn farming* (US$20 million in exports to France in 2000 plus considerable local consumption)
- PNG has significant activity in *prawn trawling*
- Some PICTs have developed niche markets in Asia for exports of *beche-de-mer*, *seaweed* and *shark fins*
- *Recreational/sports fishing* is a major tourism earner for a number of PICTs
- Activities such as *clam farming* and *live fish (aquarium) exports* have been encouraged in a number of countries and will no doubt continue to appear from time to time

A further issue to consider here is whether using different methods to catch the same type of seafood represents different activities. From discussions with a range of experienced fisheries scientists and analysts in the region there is no doubt that they see a clear need to collect separate data for different methods. But ISIC does not usually distinguish between modes of production eg, the wording "whether the work is performed by power-driven machinery or by hand" is used in the definition of “manufacturing”.

The draft classification shown below for regional fisheries studies does distinguish between methods and gear type for major activities. It also effectively differentiates between activities according to location, as the same basic activities are carried out in quite different ways in different areas eg, any fishing for tuna in coastal waters is generally very different from deep-sea tuna fishing. The three areas which have been commonly suggested for separate classification are:

- oceanic
- coastal/reef/in-shore
- inland

This effectively represents the restoration and minor extension of the location criteria that were used to split fishing in the previous version of ISIC.

In discussions with a range of fisheries experts some have argued that the three areas described are self-evident, others that it is not possible to clearly define the boundaries between them. In a regional context it may be that it is possible only to provide a broad description and that the boundaries between the three areas for any particular PICT would have to be determined on the basis of local conditions eg, the depth and size of the lagoon, the rate of drop-off outside the reef. The issue is raised here for the sake of completeness, but leaves any detailed recommendations to come from future discussions in the region.

The draft classification proposed for fishing and related activities is as follows:

05100 Coastal/reef/inshore fishing (excluding aquaculture activities)
- 05101 Trolling
- 05102 Bottom fishing in depths over 100m
- 05103 Other line fishing
- 05104 Netting
- 05105 Diving/spearfishing
- 05106 Fishing by destructive techniques (eg, dynamite, poison – this is included for completeness, but in any survey it would almost certainly be collected as part of 05107 below)
- 05107 Fishing by other methods (including fish traps)
- 05108 Collection of live aquarium fish
- 05109 Catching/trapping lobsters
- 05110 Gathering shellfish
- 05111 Gathering beche-de-mer
- 05112 Gathering clams
- 05113 Other coastal/reef/inshore fisheries-related activities

05200 Oceanic fishing
- 05201 Long lining
- 05202 Vertical long lining
- 05203 Purse seining

05204 Trolling
05205 Pole and line
05206 Prawn trawling
05207 Other fishing

05300 Inland fishing

05400 Aquaculture
05401 Pearl farming
05402 Prawn farming
05403 Other aquaculture

05500 Other marine activities

The draft classification effectively breaks the existing ISIC 0500 into five classes, three of which have a total of 23 sub-classes.

4.3 Other classifications for agriculture and fishing in the Pacific

The discussion above has focused on classifying agriculture and fishing activities according to criteria based on final outputs, methods and location. It may also be worth considering a further classification according to the underlying purpose of the activity. The possibilities which have been raised most commonly are:

- Commercial
- Artisinal (mainly for sale)
- Artisinal/subsistence (mainly for own consumption)
- Recreational

However, while there is some keen interest in having information for certain activities broken up by purpose there are very major practical problems in defining clear boundaries between the descriptions suggested above. As for the different areas for fishing suggested above, the topic is raised here for completeness and in the hope that it will help stimulate further debate.

5. Conclusion

Agriculture and fishing studies in the Pacific need to provide data at a level of detail that will support key national and intra-regional analysis. But existing international standards are more aggregated in these areas than is appropriate for the Pacific, and studies in the region have often developed their own classifications in response. Further, the international trend towards more sophisticated service industries such as electronic commerce is likely to lead to increased numbers of identified "industries" in these areas, with corresponding pressure to collapse the classification of industries with relatively small contributions to global value added. Rather than attempt to turn back the tide of ISIC development, it is proposed that a classification of agriculture and fishing activities of importance to the Pacific be developed and circulated. Hopefully this will lead to an agreed "Pacific region" standard classification that will assist survey designers and survey statisticians who are looking for guidance in these areas. This could well be a first step towards the development of a Pacific-oriented version of the full ISIC.

References

Australian Bureau of Statistics, *Quarterly estimates of national income and expenditure, (Cat No 5206.0),* December Quarter 2000

Gillett, McCoy, Rodwell and Tamate, *Tuna: A key economic resource in the Pacific Islands*, Report prepared for the Asian Development Bank and the Forum Fisheries Agency, November 2000

Secretariat of the Pacific Community, *An assessment of the role of women within fishing communities in the Republic of Palau*, 1999

Secretariat of the Pacific Community, *Selected Pacific Economies: Statistical Summary, 2000*

Treasury Department, Samoa, *Quarterly economic review*, *December quarter 2000*

United Nations Statistical Commission*, International Standard Industrial Classification of All Economic Activities (ISIC), Revision 3*

How Do Classification Systems Respond to Changing Demands of Agriculture Statistics in the New Millenium

Paul Johanis
Statistics Canada, Jean Talon Building, Section A-7, Ottawa K1A 0T6, Ontario, Canada
Tel.:+1 - 613 951 8577
e-mail: Paul.Johanis@statcan.ca

Abstract: This paper discusses current developments in the field of economic classifications and relates them to agriculture as an economic activity at the dawn of the 21st century. It describes how economic classifications can respond to emergent public policy issues such as organic agriculture, the pressures on small agricultural holdings, genetically modified organisms, agri-food and life sciences. It concludes by outlining specific changes having been implemented or being considered in the North American Industry Classification and its relationship to NACE and ISIC.

Keywords: Economic classifications, production process, NAICS

1. Introduction

Only in the Eternal City would I have accepted to give a paper with this title, which seems a bit grand and overdone. Yet, it provides an opportunity to stand back and address this topic from a global perspective, with a historical sweep not often considered in the papers tabled at such gatherings.

This paper discusses current developments in the field of economic classifications and relates them to agriculture as an economic activity at the dawn of the 21st century. It describes how economic classifications can respond to emergent public policy issues such as organic agriculture, the pressures on small agricultural holdings, genetically modified organisms, agri-food and the life sciences economy. It concludes by outlining specific changes having been implemented or being considered in the North American Industry Classification and its relationship to NACE and ISIC.

As the paper concerns the treatment of agriculture in international economic classifications, its object of interest is agriculture on the planet Earth, not only as it is practiced today in industrialized nations but as it ever was practiced somewhere in times past. Indeed, agriculture as it was ever practiced in human history is still practiced somewhere on the globe today. A classification of agriculture should therefore encompass all the various forms that agriculture as an economic activity can take.

2. An approach to the classification of agriculture: the production process

Our job in the classification community is to turn something as varied and complex as agriculture into something flat and two-dimensional, for which I apologize in advance to the non-practitioners among us. To classify economic activities, it is necessary to adopt a principle according to which like activities are grouped together. An idea that is currently making the rounds in classification circles is that economic activity should be classified according to production process. This means grouping together activities that involve similar capital and labour inputs and similar technologies by which they are combined to produce output. This is the principle that was adopted in the development of the North American Industry Classification System (NAICS) and it guided the definition of classes at all levels of the classification. Industry classifications used in North America previous to NAICS, and most industry classifications currently in use around the world, tended to group production units according to what was produced, giving little or no importance to how it was produced. The measure of a well-defined class in such systems is the coverage ratio, that is, the share of the total output of a given product in the economy accounted for by the industry class so defined. In a production process approach, this measure is secondary to the specialization ratio, which measures the homogeneity of classes in terms of their production functions.

In application to agriculture, this view could perhaps take the following form. Our Neolithic forebears invented very simple, yet effective, processes for growing crops and raising animals, which still persist today in various quarters of the world. In time, more elaborate capital equipment was introduced, along with new sources of power to supplement human effort: wind, water and draught animals. From a production process point of view, this new form of agriculture is sufficiently differentiated from its elemental predecessor to represent a new industry within the broad family of agriculture. This traditional form of agriculture then gave way to an early-mechanized form, in which first steam, and then the internal combustion engine dominate, fundamentally altering the capital-labour ratio and the economics of agriculture. Much smaller social groups could as a result efficiently handle much larger holdings, giving rise to what we know now as the family farm. Perhaps, we are witnessing yet another transformation in the production process of agriculture, with intensive large-scale mechanization emerging as the predominant form.

From a classification perspective, these four stages in the evolution of the production process of agriculture could represent four basic groupings, which, for the purposes of this paper could be labeled elemental agriculture, traditional agriculture, mechanized agriculture and intensively mechanized agriculture.

None of the currently used industry classification systems have rigorously adopted this approach. Rather, examination of ISIC, NACE and NAICS reveals that beyond the distinction between crops and animal production, priority is given to the commodity produced in defining agricultural industries. For example, the wheat growing industry is a class in each of these classifications, regardless of the actual production process used. Small, hand-tended holdings are classified alongside immense, intensively mechanized operations. This would make sense if the only matter of interest were the total production of wheat but from the point of view of productivity analysis, industry structure and many of the policy issues in the field of agriculture today, perhaps a supply-side approach would be more effective. Total supply of a given commodity would be more accurately estimated by collecting commodity data in production surveys, using appropriate commodity classifications, rather than by using industry classes as a proxy for commodity. It is the matrix produced by combining industry and commodity classifications that provides a complete picture of economic activity, with the use of inputs described in the industry dimension and the supply of outputs described in the commodity dimension. For the purposes of this paper, let us assume that such an approach can be adopted for agriculture. In what way could this help in responding to public policy issues in agriculture today?

3. Issues

3.1 Boundary issues

It may seem that we can therefore place agriculture neatly in a box defined by production process on one axis and commodity on the other. Alas, all is not so clear. As much as classification practitioners would like to see the world in black and white, they must contend with multifarious shades of gray. Two major boundary issues have bedeviled the classification of agriculture in national and international classifications for some time: 1) the differentiation between subsistence activities and agriculture; and, 2) the distinction between agriculture and manufacturing. From a production process point of view, we might say where does agricultural activity begin and where does it end.

The first boundary issue is really one of specialization. Production units engaged in subsistence activities, usually households, almost by definition engage in a variety of undifferentiated economic activities, each of which could be classified on its own: agriculture, hunting, construction, making clothes, tending for family members etc. The difficulty is in classifying units engaged in all of these activities, in different proportions at different times. It is to deal with this issue that ISIC rev. 3.1, which will be published in 2002, includes a proposal for two new classes: undifferentiated goods-producing activities of households for own use; and, undifferentiated services-producing activities of households for own use. This will provide a more definite boundary at the front end of agriculture.

The second boundary issue concerns the distinction between agriculture and manufacturing. This issue can be decomposed into two interrelated parts: one dealing with vertical integration, the other turning on the definition of an activity called beneficiation.

Some production units engaged in agricultural activity are also engaged in preparing agricultural products for further transformation and, in some cases, in transforming these products into food and other products. The classification of vertically integrated units, where the output of one process is the input to another process within the same production unit, always poses problems and requires that conventions or standard rules be adopted. For the most part, industry classification systems require that vertically integrated activities be classified to the last activity in the integrated production process, with some exceptions. Where an upstream activity is overwhelmingly dominant in terms of its contribution to the value-added in the product, then the whole integrated process can be classified to the upstream activity. An example of this exception is the treatment in NAICS of the integrated production of steel and steel forms. Such integrated units are classified to steel mills rather than to the fabricated metal product industry, as would be required by the vertical integration rule. In other cases, exceptions to this rule are made for historical, institutional or other reasons, as in the case of integrated grape growing and wine production on agricultural estates, which is classified in ISIC to agriculture. Such exceptions are often not explicitly acknowledged and vary among jurisdictions. This obscures the boundary between agriculture and manufacturing.

Even if all jurisdictions agreed on the list of exceptions to the vertical integration rule and their treatment, there would remain a difficulty in establishing the "back-end" boundary of agriculture. This concerns the treatment of an activity known as beneficiation. The products of primary industries often require some kind of pre-manufacturing treatment to make them transportable or otherwise processed. In the case of agricultural products, this can involve cleaning, sorting, de-stemming, drying and other activities, depending on the crop. While it is recognized that integrated agricultural and beneficiation activities are to be classified to agriculture, the classification of production units primarily engaged in beneficiation only is problematic. The prototypical case is cotton ginning, which is considered by some as an agricultural activity, by others as a manufacturing activity. At issue here is the definition of beneficiation. At what point does an activity that prepares a product for further processing cease to be an agricultural activity and become a manufacturing one? Various attempts have been made to define a general principle that could be used to define beneficiation for all types of products. They often invoke transportability or access to primary markets as factors in the definition. However, markets are organized in different ways in different countries so that it is not always clear what the primary market is for agricultural products, for example, in the case where the crop is sold even before it is put into the ground. In the end, perhaps the only way of establishing a clearer boundary between agriculture and manufacturing will be to itemize beneficiation activities for each type of crop and ruling case by case on its proper classification. In any event, it seems that this boundary will remain somewhat smudgy for some time to come.

3.2 Public policy issues

Assuming that one can construct a classification system for agriculture that takes into account fundamental production process differences, how could this help illuminate public policy issues in the field of agriculture and inform public debate? Five current policy issues will be examined from this perspective: organic agriculture, the family farm, GMOs, agri-food and the life sciences economy.

3.2.1 Organic agriculture

In many jurisdictions, agriculture using organic methods is moving from the fringe to the mainstream. Some countries (for example, Austria) have adopted specific targets for increasing the share of organic agriculture in overall agricultural production. Consumer concerns for security of food supply and environmental consciousness are factors in the rising popularity of organically grown products, and healthy profit margins are a boon to producers. To track the evolution of organic agriculture, an industry classification based on production process would be a valuable instrument. Although we speak of organic products, these are often indistinguishable from "non-organic" products. The real difference lies in the production process. The hallmarks of organic farming are its methods, such as natural pest control, crop-diversification and use of manure and compost as fertilizers. This may represent a sufficiently different production function (capital, labour, energy, materials and services) as to define a different industry within agriculture. Somewhere between mechanized agriculture and intensively mechanized agriculture, a sub-class called organic agriculture could be defined. The evolution of the number of production units classified to this class according to their production process would be an important policy variable.

3.2.2 Pressure on the family farm

Another phenomenon of current policy interest that might be reflected in a classification system based on production process is the plight of the family farm where institutional frameworks and market structures no longer sustain small, family-run agricultural holdings. While there may be many economic and social factors involved in the difficulties faced by family farm operators, technological factors also surely play a role. From a purely production process perspective, we may characterize this change as a transition from mechanized to intensively mechanized agriculture. Again, the evolution of the number of units classified according to this production process would be an important policy input.

3.2.3 Genetically modified organisms

The use of genetically modified organisms in agriculture raises a multitude of concerns and possibilities for consumers and producers alike. Governments, researchers and other data users have an expressed need for tracking the extent to which GMOs are used in agricultural production. The question is whether a classification system based on production process would be of any use in this respect. For the most part, GMO crops are not grown differently from the non-GMO crops. The production process is the same, except for the seed and maybe some additional labeling and record keeping. This could not form the basis, however, for the delineation of a separate "GMO agriculture" industry. Rather, it would seem that this issue would be better addressed in the product dimension, with the addition of GMO product codes alongside their non-GMO counterparts. Producers are required for the most part to keep this information in their records, and could therefore supply it for statistical purposes. In addition, consumers will likely exert increasing pressure to maintain this distinction throughout the supply chain, so that manufacturers and distributors would also be required to keep this information. As most of these actors would likely be involved in the production and distribution of both GMO and non-GMO products, an industry classification system based on production process would be of no assistance in tracking this phenomenon. An expanded product classification would be the better instrument.

3.2.4 Agri-food

Many public policy issues in this field, from security of the food system, to health of the environment and innovation for competitiveness and growth, do not concern agriculture alone but a broader industrial supply chain that includes manufacturing and distribution. This composite industrial complex is sometimes referred to as the agri-food sector. In response to this expanded policy focus, the Canadian Department of Agriculture broadened its mandate and changed its name to Agriculture and Agri-food Canada (AAFC) nearly a decade ago. From a classification standpoint, the agri-food sector can be defined as a special aggregation within an industry classification system, grouping together classes from the agriculture, manufacturing and distribution sectors. Problems arise when classes in the non-agricultural sectors have not been defined with food as a primary defining factor. This is usually not a problem in manufacturing, where most classification systems have defined food manufacturing industries, but classes in wholesale and retail are often too general for this purpose. Future revisions of industry classification systems should therefore take into consideration the special aggregations that will likely be required to support policy and economic analysis. In the case of agri-food, while a fairly clean special aggregation can now be defined under most national and international industry classification systems, changes can be made in the next rounds of revisions to improve the ease with which this special aggregation can be made.

3.2.5 Life science economy

The life science economy, which may be updated terminology for the set of economic activities that is also referred to as the biotechnology sector, is an even broader industrial complex. In addition to food products, agricultural products can also be used to produce bio-fuels such as ethanol, or methane from biomass. Agriculture is also a source for nutraceuticals, bio-pharmaceuticals, building materials, plastics and paper. AAFC speaks of the life science economy as the use of knowledge of living things to create new bio-based products. Although statistical agencies have very often been called upon to define this sector as a special aggregation, the creation of a special aggregation for such a broad and generally defined set of activities has proven problematic. Often, relevant manufacturing and distribution classes have not been defined along organic and inorganic lines. There is also a large research and professional services component in scope of this special aggregation, which also have often not been defined in terms of life sciences versus other sciences. Finally, it is not clear whether to include in the special aggregation all industries that use bio-products and processes or only those that produce them. At this point, industry classifications, whether based on production process or otherwise, are not well suited to define the life science economy.

4. Conclusion

This paper has so far mused upon the possibilities, use and relevance of an industry classification system that is based on the production process principle as applied to agriculture. In actual fact, however, no current industry classification has fully implemented this approach nor have any concrete plans been laid out to do so. We are, however, just beginning consultations and research for the next round of revisions in 2007, and this paper may generate ideas for proposals along the lines described above. The next round of revisions to industry classifications is significant because the calendar of revisions for ISIC, and therefore NACE, and NAICS all coincide in 2007. This may present an opportunity to achieve greater convergence among these classifications.

While no industry classification has fully implemented the approach described in this paper, NAICS 1997 did in fact introduce a number of changes in the definition of the agriculture sector based on the production process principle. Within the crops subsector, a production process distinction is made between field crops and horticulture, and within the latter, a further distinction is made for horticulture crops grown under cover. These are modest changes, but they reflect a production process approach. In addition, activities that were classified in agriculture in the previous industry classification system have been moved to other sectors in NAICS, as a result of the production process principle. These include operating irrigation systems (now in Utilities), veterinary services (now in Professional, Scientific and Technical Services) and pet grooming services (now in Other Services). In NAICS 2002, there is one minor change the agriculture sector, the addition of a Canadian industry for combined egg and poultry production.

In ISIC Rev. 3.1, which will be published in 2002, the main change that relates to agriculture is the introduction of two new classes for the undifferentiated subsistence activities of households, both goods-producing and services producing, which will help clarify the boundary between subsistence activities and agriculture. There is also an expressed intention to publish more special aggregations as appendices to the manual and agri-food is under consideration as a candidate.

As the next round of revisions to ISIC, NACE and NAICS is planned for 2007, the statistical agencies responsible for these classifications have undertaken a feasibility assessment of achieving greater convergence between them. Convergence can take two fundamental forms: the adoption of the same structure or an improvement over existing concordances between ISIC, NACE and NAICS. In addition, each of these approaches can be applied at various levels of detail of the classifications: at the top level only; at the most detailed level (defined for this purpose as the level at which the 3-country NAICS is harmonized, representing 478 classes) or at some middle level. In the case of agriculture, the working group is assessing a scenario in which the top three levels of the classification would be the same in ISIC, NACE and NAICS. A full report on convergence will be made to the Statistical Commission of the United Nations next spring.

Regardless of the outcome of the convergence discussions, a call for proposals to update ISIC in 2007 will soon be made. It is through this process that ideas and suggestions will be solicited to ensure that classification systems do respond to changing demands of agriculture statistics in the new millennium and I hope that this paper will have made a modest contribution in this endeavor.

References

Statistics Canada, North American Industry Classification System, Canada, 1997, catalogue number 12-501-XPE

Agriculture and Agri-Food Canada, Report on Plans and Priorities 2001-02, www.canada.gc.ca

United Nations, International Standard Classification of All Economic Activities (1990), Third Revision, United Nations Statistical Papers, Series M, No. 4, Rev.3, New York.

Eurostat, Nomenclature générale des activités économiques dans les Communautés Européennes (NACE), Rev.1, 1993

6 June 2001

Invited Paper Session

AGRICULTURAL SURVEY METHODS IN TRANSITION AND DEVELOPING COUNTRIES

Chair: Á. González Villalobos

Agricultural Statistics in Iran

Abbas Gerami
Head of the Agricultural Statistics and Information Department,
Ministry of Agriculture of Iran
e-mail: gerami@asid.moa.or.ir

Abstract: This paper presents an outline of the Agricultural Statistical System of the Islamic Republic of Iran, including the Agricultural Census, conducted since 1973**.**

Keywords: Agricultural surveys, national statistical systems**.**

1. General characteristics of the country

1.1 Geographical characteristics and administrative divisions

Iran comprises a land area of over 1.6 million square km. It lies at the northern temperate zone, between latitudes 25° 00′ N and 39° 47′ N and between longitudes 44° 02′ E. The land's average height is over 1200 meters. The lowest place, located in Chalet-loot is only 56 meters high, while the highest point, Damavand peak in Alborz mountains, rises as high as 5610 meters. At some points of the southern coastal strip of the Caspian Sea, land height is 28 meters lower than open seas.
Iran is bounded by Turkmenistan, the Caspian Sea, Azerbaijan and Armenia on the north, Afghanistan and Pakistan on the east, Oman Sea and Persian Gulf on the south, and Iraq and Turkey on the west.
Administratively the country is divided into Ostans (provinces), Ostans into Shahrestans (cities), Shahrestans into Dehestans (districts). According to the most recent administrative division the country is divided into 28 Ostans, 282 Shahrestans, and 724 Dehestans.

1.2 Climate

The country has three climatic zones:

A. Arid/semi-arid regions of interior and far south which are characterized by long, warm and dry periods, lasting sometimes over 7 months. The annual precipitation is such regions vary between 30 to 250 mm.

B. Mountainous extensions which are in turn divided into cold and moderate mountainous regions:

- Cold mountainous regions: About 40,000 sq.km of the total land consists of major highlands, including Alborz and Zagross mountain ranges, as well as Sahand and Sabalan high peaks. The annual precipitation in these regions exceeds 500 mm.

- Moderate mountainous regions: some 300,000 sq. km of the country enjoys moderate mountainous climate, where the annual precipitation varies from 250 to 600 mm.

C. Caspian region, which is a narrow strip with a limited extension, trapped between the Caspian Sea and Alborz Mountains, with as much as 600 to 2000 mm precipitation per annum.

2. Importance of the Agricultural Sector

- Nearly 27.4 percent of the population are engaged in agricultural activities.
- About 25 percent of Gross Domestic Product (GDP) are derived from agriculture.
- Produces more than 80 percent of food.
- More than 25 percent of export (except oil) earnings are from agriculture.
- Source of 90 percent of raw material of Agro-industries.
- Potential sector for reducing unemployment.

In view of the importance of agriculture sector, the availability of reliable and timely agricultural statistics becomes extremely important for planning by the policy makers, agricultural producers, traders, consumers, exporters and importers for effective planning and management process whether it relates to production, marketing, consumption, processing, export or import.

3. Organization of Agricultural Statistics in Iran

In Iran, agricultural statistics activities are conducted under a decentralized system that involves different governmental departments undertaking statistical surveys in their areas of responsibility. There is a High Commission on Statistics, which coordinates these activities.
The major departments involved in agricultural statistics in Iran are as following:

- *The Statistical Center of Iran*
- *The Agricultural Statistics & Information Department of the Ministry of Agriculture*
- *The Ministry of Jahad (Fisheries Forestry and Livestock)*

3.1 Statistical Centre of Iran

The *Statistical Centre of Iran* was established in 1965 and is one of the government agencies with a departmental status under the office of Government organization for planning and management. For the time being the centre is responsible for conducting all censuses including the agricultural censuses.

3.2 Agricultural Statistics & Information Department of the Ministry of Agriculture

The Agricultural Statistics and Information Department (ASID), established in 1981 under the secretary of planning and budget is the only agency of the Ministry of Agriculture (MOA) responsible for the collection, compilation and publication of all Current Agricultural Statistics.

3.3 Ministry of Jahad

Until 1978, the *Statistical Centre of Iran* and the *Agricultural Statistics & Information Department of the Ministry of Agriculture* were the only two governmental agencies responsible for agricultural surveys. After 1978 the *Ministry of Jahad* has also been involved in some agricultural surveys mainly in the field of livestock, fisheries and forestry.

4. The Agricultural Censuses in Iran

The Agricultural Censuses in Iran have been conducted by the Statistical Centre of Iran.

- The first Agriculture Census was conducted in 1973.
- The second Census of Agriculture was conducted in 1988 in urban and rural areas by visiting settled and unsettled households to collect data on all the agricultural holdings belonging to physical persons as well as data on agricultural holdings belonging to legal entities by identifying all places used by them for agricultural activities.
- The third Census of Agriculture was carried out in 1993. In the census, some cities or city districts, and some holdings which were not significant in the 1988 Census, were excluded.
- The fourth census of Agriculture is scheduled to be conducted in 2003.

5. The Current Agricultural Surveys

The General Department of Public Statistics conducted the first National Agricultural Survey in 1960. Since then, the following surveys have been carried out:

5.1. Current Agricultural Sample Surveys conducted by the Statistical Centre of Iran

Since its establishment, the Statistical Centre of Iran was involved in different statistical activities, and it carried out the following agricultural sample surveys:

a) Agricultural Sample Survey of holdings, investigationg farming activities, production, value of products and production costs of major crops (Conducted in 1967, 1971, 1972 and 1992)

b) Survey of modern cattle farms (conducted in 1988, 1993 and 1995)

c) Chicken farm survey (conducted in 1985 and 1986)

d) Livestock Surveys -slaughtered data in slaughterhouses- (conducted yearly since1968 until now).

5.2 Current Agricultural Sample Surveys conducted by the Agricultural Statistics & Information Department of the Ministry of Agriculture

Before establishment of the Agricultural Statistics and Information Department, the *Economic and Statistics Department* of the Ministry of Agriculture carried out some surveys.

5.2.1 The four Main Regular Yearly Probability Sample Surveys

Crop	Sampled Villages	Sampled farmers
1. Wheat and Barley	5737 (≈8%)	20-30 percent
2. Rice	998 (≈6%)	20-30 percent
3. Other annual Crops	5419 (≈8%)	20-30 percent
4. Cost production*		

* The sample consists of a 10 to 20 percent sabsample of farmers selected from the sample used for the other three surveys

Survey design for the Wheat and Barley Survey, the Rice Survey and the Other Annual Crops Survey

These three surveys use a common two-stage systematic stratified sample design, and a common sample:
For the first sampling stage, villages are stratified into 5 to 7 strata based on some relevant auxiliary variables from the latest agricultural census. The total sample size (n) is determined by two factors, namely, required accuracy and costs. The optimal (Neyman) allocation is used to assign sample size to each stratum. Systematic sampling is used to select villages within each stratum.
For the second sampling stage, a list of farm households is prepared for each selected Village. Then 20-30 percent of farm households are randomly selected from the list.
The three surveys provide estimates at city level. All sample farmers are interviewed by using multipurpose questionnaires, and direct interviews with farmers.

Sample design for the Cost of Production Survey

For the Cost Production Survey, the sample consists of a subsample of farm households from the previously selected sample of farm households used in the previously mentioned surveys.
The survey provides estimates at provincial level. All sample farmers are interviewed by using multipurpose questionnaires, and direct interviews with farmers.

5.2.2 Survey of Retail and Wholesale Agricultural Market Prices

These Surveys are conducted at provincial level every two weeks. The results of these surveys are disseminated at provincial level.

5.2.3 Ad-hoc Agricultural Surveys

The following Ad-hoc Agricultural Surveys have been conducted:

A. Probability Surveys to estimate crop area of sugar beet in certain provinces, in collaboration with the relevant Departments

B. Probability Surveys to estimate crop area and production of potato and onion in almost all provinces, in collaboration with the relevant Departments

C. Crop area estimation for some crops in some provinces using estimates derived from satellite data

D. Crop yield forecasting for some crops in one province using estimates derived from satellite data

5.2.4 Yearly Crop Forecasting Surveys

For major crops such as wheat, rice, yearly production forecasting is made through a reporting system prior to harvesting time.

5.3 Recent Current Agricultural Surveys methodological developments

In 1995 a research contract was signed with the International Institute for Aerospace Survey and Earth Sciences (ITC) to design and develop a proper method which can generate timely and reliable information on the area and production of the major agricultural commodities in the Hamadan province, which could be eventually extended to cover the entire national territory. As far as the area of agricultural commodities is concerned the result seems satisfactory. But this is not the case with production forecasting and estimation. A follow up project with ITC is being considered to find models to obtain a more accurate estimation and prediction.
Several aspects of the current survey designs are under revision:

Using ratio and regression estimators
Finding better classifier variables
Using cluster analysis methodologies as a better method to classify villages, rather than the Dalenius method that is currently being used.
Utilizing replicated sampling.

In 2000 ASID request a mission from FAO to advise on the establishment of a new national programme of current agricultural surveys based on a statistical survey model using multiple frame methods. A proposal has been prepared for approval.
The data processing system is computerized. Besides the release of the results in printed format, efforts have been made to make them available through magnetic media, such as FD and CD-ROM, facsimile communications and internet. The homepage of ASID is http://www.moa.or.ir

5.4 The Agricultural Statistics & Information Department organizational structure

The Agricultural Statistics & Information Department has a central and provincial agencies, as shown below:

Central Organization ↔	Provincial level (29 Offices)
MOA, ASID	Provincial Agricultural Organization Bureau of Statistics

The provincial offices in addition to implementing the survey and compiling the results for their prefectures, provide instruction to their prefectural offices under their jurisdiction, and report to the central office (ASID).

5.5 Current Agricultural Sample Surveys conducted by the Ministry of Jahad

Before 1978, only the Statistical Centre of Iran and the Agricultural Statistics & Information Department of the Ministry of Agriculture were involved in collecting agricultural statistics. Since 1978, the Ministry of Jahad has been also involved in some agricultural surveys, such as surveys on livestock, fisheries and forestry as well as on development of rural areas.
The current agricultural surveys conducted are the following:

- Statistical survey of fish breeding centres for fish farms
- Statistical survey of fish farms
- Farm survey
- Statistical survey of cheese production units
- Statistical survey of natural and semi-natural water resources
- Statistical survey of broiler chicken farms
- Statistical survey of modern cattle farms

- Statistical survey of breeder chicken farms
- Statistical survey of hatchery units

Bibliography:

Kheiri,G.(1998). *A Report on Statistical Activities in Agriculture, Forestry and Fisheries in Iran*. Persented at the Seminar on Statistics in Agriculture, Forestry and Fisheries in Iran, 21-30 July 1998.

Agricultural Statistics and Information Department (2000*)*. *Current Activities in the Year 1999*. A Report to the Deputy Ministry of Agriculture in Planning and Budget.

Statistical Centre of Iran (2001). *Iran Statistical Yearbook*.

Gonzalez Villalobos, A (2000). *FAO Report to Iran to Advise on the Establishment of a National Programme of Current Agricultural Surveys*.

The Agricultural Census in Hungary: Analysis of Results

Éva Laczka
Hungarian Central Statistical Office, Agricultural Statistics Department,
Head of Department,
H-1024, Budapest, Hungary, Keleti Károly u. 5-7,
E-mail: eva.laczka@office.ksh.hu

Abstract: The first Hungarian Agricultural Census took place in 1895. After then similar surveys were carried out in 1935, 1972, 1981 and 1991. The Agricultural Census 2000 was the sixth in the series of agricultural censuses. The aim of the 2000 census was to survey the economic structure created after the reprivatization of the landed properties, but the EU and FAO requirements will also be observed at the shaping up of the programme.
The paper gives an overview on the implementation of the Agricultural Census (legal background, role of the task forces, design of farms, completeness, division of labour between the Department of Agriculture and HCSO regional directorates, communication)
Concerning the result, the main characteristics of the Hungarian Agriculture are highlighted (huge number of small family farms). The list of publications is also included.

Keywords: number of holdings, history, preparatory phase, completeness, training of enumerators, communication, results, characteristics, publications

1. Hungarian Agriculture

Agriculture has played a significant role in our country's economy in the last century and the same can be said even today. The cultivated agricultural land area is 70% of the total land area, while 7% of the active earners are employed in agriculture. The share of agriculture in the Gross Domestic Products (GDP) is about 5%, and that of the food industry is more than 6%.
In the recent years fundamental changes have taken place as regards the ownership and the structure of agriculture in Hungary. In previous years a relatively small number of large-scale agricultural and food-industrial holdings existed. After the privatisation of the land area, state farms and the transition of agricultural co-operatives, a large number of small and medium size agricultural units were established. According to data of the Agricultural Census 2000, there were more than 8000 companies, state farms, co-operatives and other agricultural enterprises in the country. Beside them 960 thousand households qualified as farms were also surveyed. (Table 1)

2. Agricultural Censuses 1895 – 2000

Regular agricultural statistical data supplying in Hungary dates back to almost two hundred years. The surveys conducted from 1828 covered the land sown, agricultural production and number of productive livestock. The first land register, the so-called provisional cadaster – providing the basis for subsequent land area statistics – was introduced in 1853 while the data collection relating to vineyards and vintages was introduced in 1873.

The first detailed agricultural census was ordered by statute VIII passed in 1895. The survey covered every owner-occupied farm.

The second census was accomplished in 1935 when it was recommended by the International Institute of Agriculture (IIA) in Rome, the predecessor of FAO.

The third census was ordered by a Cabinet decision No 3401/1970 and a governmental decree No 2/1972 regulated its execution. In compliance with the government's decision the census took place in 1972.

In 1980 Hungary joined the world census indicating to FAO applying of cost-saving methods which implied narrowing down the data collection of 1972.

Taking into account the data requirements of national institutions and FAO every five years, the HCSO, in conjunction with the livestock census in 1976 and 1986, collected information also on the most important data of small-scale producers.

The subsequent agricultural census in 1991 was ordered by a government decision dated 2 August 1990 and the execution of the census was regulated by decree No 36/1991 of 1 March 1991. The objectives of the census were realized by the harmonized accomplishment of several surveys.

3. The Sixth Agricultural Census

Pursuant to Act XLVI of 1999 the Hungarian Central Statistical Office (HCSO) carried out an Agricultural Census (AC) between 1. – 21. April 2000 by the reference date of 31. March 2000. The AC 2000 was the sixth in the series of agricultural censuses.

3.1. Preparatory phase

January 5, 1999	The AC plan was accepted by the HCSO Presidency
January 22, 1999	The act on AC was drafted
February, 1999	Interdepartmental review of the act
February 10, 1999	Governmental decision on AC
March and April, 1999	Review of the act by parliamentary committees
May 4, 1999	The act on AC was passed by the Parliament

All documents adjunct to AC have been reviewed by professional bodies, including the Ministry of Agriculture and Regional Development (MARD), professional associations and unions, universities and research institutes, and the regional directorates of HCSO. All comments, specifically those of MARD, have been included in the final design of the questionnaire, provided there was no conflict with the relevant EU regulations and FAO recommendations.

A Professional Committee for the coordination of methodological and organizational tasks, and a Financial and Budgetary Committee for preparing the financial transactions were created. The participants have been delegated by the MARD, regional directorates and the different departments of HCSO, including the department of Agriculture. The task forces held regular and ad-hoc meetings.

3.2. Completeness – Interview districts

One of the most difficult tasks in the design of censuses is the assurance of completeness. Apart from the family farms agricultural businesses need the same foresight for circumspect mapping. The changes in the past years and the not up to date registry added to the difficulties in the survey of businesses. For managing this issue a new report was introduced.

Similarly to the earlier censuses, completeness on family holders was achieved by visiting each household in the districts of settlements where animal husbandry was permitted. Apart from the rural settlements interview districts were also created in towns during the year 2000. For arranging the interview districts administrative sources, MARD registry data were used and the registries of farmers were consulted, too.

A total of 13,712 interview districts were selected in the light of local peculiarities with a view to complete the census in each district in three weeks time at most. The regional directorates and the local authorities defined jointly the boundaries of interview districts to include in the census all inhabited parts of each settlement without any overlapping. The interview districts were determined by the end of January 2000.

3.3. Training of enumerators

Between February 28 and March 2, 2000 four full-day briefing sessions were held in Budapest for the staff of HCSO regional directorates (the number of them is 19) involved in the census, including the county census officers and his/her deputies, and the census coordinators. The county staff was given first-hand information about the tasks involved in the implementation of the census, and had a chance to clarify issues and questions.

Qualified regional officers of the regional directorates trained the enumerators, supervisors and settlement officers on the basis of the implementation directive of January 31, 2000 and the training curriculum of February 28, 2000 issued by the department of Agriculture.

3.4. Communication

A Communications Ltd. designed the AC communications programme under contract, in cooperation with the staff of the Department of Agriculture.

Two press conferences were held before the census (08. 12. 1999 and 21. 03. 2000). The preliminary data of AC were already presented in June 2000, the final data of AC were presented in September 2000 at press conferences.

3.5. Results

In the framework of the census, the enumerators have visited 2,1 million households (more than 60% of total households). 960,000 households have reached the reporting holding size. Another 835,000 households control smaller land area or livestock than the specified threshold, and 300,000 households do not perform any agricultural activities.
According to the data, on 31. March 2000 the agricultural sector in Hungary includes nearly 960,000 family holdings and more than 8,400 enterprises engaged in agricultural activities. In the census completed in 1991 2,600 enterprises and 1,4 million family holdings were recorded.
27% of the enterprises do not cultivate any land area. More than 70% of the family holdings utilise less than 1 ha or altogether they farm less than 8% of the total productive land area used by the group of family holdings.
The structure of the livestock is dominated by cattle in case of enterprises and by pigs in case of family holdings. 90% of the total livestock consists of 2 species in case of enterprises and 4 species in case of family holdings.
From the total livestock of family holdings more than 40% of cattle, nearly 50% of cows, 36% of pigs, and 41% of breeding sows are kept by the farmers controlling 1 to 10 ha of productive land area. This sector controls more than 30% of the cow- and pig-keeping capacity and nearly 50% of the wine storage capacity. This sector has a considerable share in the stock of tractors and trucks, too (42% and 38%). In case of the enterprises, these resources are concentrated in units using 100 ha or more.
Slightly below 3 persons on the average belongs to a family holding, 2 of them actively participate in farming.
Nearly 25% of holders are females. The average age of the holders is 60 in case of females and 53 in case of males. (Tables 2-3.)

3.6. Some characteristics

Due to the large number of petty farms producing exclusively for family consumption, compliance with the statistical coverage specification was an extremely difficult methodological task. At the same time, comprehensive monitoring of agricultural activity provides vital information for decision-makers, analysts and agents of the market. For instance, the livestock kept at market producer farms is insufficient for determining the total livestock. The exact definition of the market producer unit would also bring up numerous problems, not mentioning comparability. In terms of size, activity or degree of specialization, commodity producer farms under Hungarian circumstances probably would not fall in the same category prevailing in the Netherlands or Denmark.
In Hungary market production of petty farms producing mainly for own consumption was not negligible even at the time of earlier censuses conducted in 1972 and 1981. The picture was the same in 2000, too.
For analyzing the structure of family farms, the gross production value of the units was estimated. The estimated gross production value was based on average prices and yields rather than the actual output of individual farms. Product balance sheets were used as the basis of calculation. At that time, gross production value was the only value

indicator for summarizing data expressed in terms of various different natural units, such as land area by land use categories, livestock by kind of animals, etc.
In our investigations, size groups of farms were defined on the basis of gross production value. Our findings showed that most of the family farms turned out to have an extremely low production value. The ratio of farms where the majority of income of a small farming household originated from agricultural production came to only a few percent on the basis of the production value net of costs. Households depending exclusively on agricultural production could be assumed only in the case of the largest farms while in the case of the majority of farms other sources of income also contributed to make a living. Part-time agricultural activity, which has become a worldwide phenomenon, is widely practised in Hungary, too. (Table 4.)

3.7. Publications

Following the census, the processing of the data started immediately, thus the first, preliminary data were already published in June by the HCSO. Publishing of the final data commenced in September.

The Agricultural Census 2000.

The CD made by our Department on this subject provides overall information about the history and organisational structure of the HCSO, emphasizing the role of the Department of Agriculture. It contains preliminary information on the recent agricultural census and a general overview on the changes taken place in the Hungarian agriculture over the last decade. It discusses the history and most important findings of the previous censuses.

Hungarian Agriculture 1851-2000

The CD gives a historical summary of the Hungarian agriculture based on the content of a previous CD. However is extended with the preliminary data collected by the Agricultural Census in spring 2000, in a separate chapter
The CD contains two analyses. One is a summary of the development of the Hungarian agriculture in the 20th century. The other gives a comprehensive look at the agricultural production, value of production and agricultural prices of the year 1999.
The long trend figures of agricultural labour force, land area by land use categories, sown area and yield production of the main crops, livestock and the livestock products, gross agricultural output, average procurement prices, and market prices are shown in tables for easy diagram calculations.

Hungarian Agriculture, 2000. - regional data -

This publication contains data of the territorial units (150 NUTS IV, 19+1 counties, 7 regions), on the number of the holdings, on the size, structure and average size of the land use, the livestock, on the machinery and equipment, buildings and structures, on the agricultural production value on holding level, on the distribution of the production value of the holdings, on the characteristics of the land use and livestock per holding concerning the family holdings and agricultural enterprises in 123 tables. Available in paper and CD format.

Land Use of Hungary, 2000. - data by settlements -

This publication shows by settlements the enumerated agricultural land area by land use categories, the number of the users, quality of arable land area, fruit tree plantations and vineyards, the territorial land within the administrative boundary, respectively provides data on family holdings and agricultural enterprises by counties and regions. Available in paper and CD format.

Livestock in Hungary, 31 March 2000. - data by settlements

This publication informs by settlements about the size and the structure of the livestock, the number of livestock keepers, respectively their rank by counties and regions. It contains county and regional data concerning the family holdings and enterprises, as well. Available in paper and CD format.

Agricultural fixed assets, 2000 – data by counties

This publication includes data on the number and the capacity of the various types of machinery and equipment, farm buildings and structures. Data on the age of assets and other technical parameters are also published.

Under preparation

Agricultural employment, 2000 – regional data

This publication includes data on the number of persons belonging to the holding, the number of persons working at the holdings, the time spent with agricultural work, the volume of paid and non-paid labour etc. It will be available in July 2001.

Table 1: *Variation of the number of holdings*

thousand

Year	Family holdings	Enterprises	Total
1972	1841,5	6,1	1847,6
1981	1529,6	1,4	1531,0
1991	1395,8	2,6	1398,3
2000	958,3	8,4	966,6

Table 2: *Main characteristics of enterprises by size of the productive land area*

percent

Denomination	Size of the productive land area					Total
	0	<1.00	1,00-<10,00	10,00-<100,00	100,00	
	hectare					
Number of enterprises	27,0	4,0	9,2	27,1	**32,6**	100,0
Productive land area	0,0	0,0	0,1	2,3	**97,6**	100,0
Cattle	5,5	2,3	0,2	2,3	**89,8**	100,0
Cows	5,5	2,0	0,2	2,5	**89,8**	100,0
Pigs	25,1	0,5	3,6	6,9	**63,9**	100,0
Breeding sows	25,7	0,4	3,7	7,0	**63,2**	100,0
Horses	12,7	0,6	3,0	16,8	**66,9**	100,0
Sheep	8,0	1,6	1,0	7,1	**82,3**	100,0
Hens, chickens, roosters	**50,7**	0,6	3,4	7,3	38,0	100,0
Geese	**61,7**	0,1	0,3	14,7	23,3	100,0
Ducks	**81,9**	0,5	4,0	9,2	4,4	100,0
Turkeys	**77,1**	0,5	0,2	5,6	16,6	100,0
Cowshed capacity	6,0	2,7	0,3	3,1	**87,8**	100,0
Pigsty capacity	24,8	0,7	3,3	6,3	**64,9**	100,0
Henhouse capacity, m^2	**51,0**	0,5	2,0	7,4	39,2	100,0
Wine storage capacity, hectolitre	2,6	0,2	1,8	9,7	**85,7**	100,0
Tractors, piece	8,5	1,8	1,5	7,4	**80,8**	100,0
Trucks, piece	13,5	1,7	2,4	7,4	**75,1**	100,0

Table 3: *Main characteristics of family farms by size of productive land area*

percent

Denomination	Size of the productive land area					Total
	0	<1.00	1,00-<10,00	10,00-<100,00	100,00	
	hectare					
Number of holding	0,4	**70,0**	24,2	5,1	0,2	100,0
Productive land area	0,0	7,7	27,7	**47,7**	16,9	100,0
Cattle	0,2	16,0	**41,6**	35,6	6,6	100,0
Cows	0,1	13,7	**48,7**	31,0	6,6	100,0
Pigs	0,6	**43,2**	36,4	17,4	2,4	100,0
Breeding sows	0,9	35,0	**41,4**	19,1	3,6	100,0
Horses	0,7	**48,5**	34,8	13,5	2,5	100,0
Sheep	0,1	12,3	21,8	**41,5**	24,3	100,0
Hens, chickens, roosters	13,6	**50,9**	22,7	12,3	0,5	100,0
Geese	0,0	24,1	**30,3**	**29,4**	16,1	100,0
Ducks	1,9	26,6	**53,8**	17,6	0,1	100,0
Turkeys	17,1	**57,9**	18,2	0,9	6,0	100,0
Cowshed capacity	0,2	22,7	**36,8**	**34,8**	5,5	100,0
Pigsty capacity	0,5	**49,1**	34,0	14,7	1,7	100,0
Henhouse capacity, square m	11,7	**49,9**	25,4	11,1	2,0	100,0
Wine storage capacity, hectolitre	0,0	36,1	**46,1**	17,2	0,6	100,0
Tractors, piece	0,1	17,0	**42,1**	**35,7**	5,1	100,0
Trucks, piece	0,3	20,3	**37,6**	**35,0**	6,9	100,0

Table 4: *Key indicators of family farming units, 2000*

	Gross production value (thousand HUF)							Total
	-50	51-100	101-200	201-500	501-1000	1001-5000	5001-	
Number of farms (thousand)	114345	196094	250285	213398	88722	83150	12390	958384
Share of farms (percent)	11,9	20,5	26,1	22,3	9,3	8,7	1,3	100,0
Share in gross production value (percent)	0,7	2,9	7,2	13,1	12,3	32,9	30,9	100,0
Per one farming unit								
arable land (hectares)	0,39	0,36	0,53	1,2	3,0	9,55	40,33	2,75
vineyard (hectares)	0,09	0,13	0,17	0,30	0,59	1,01	2,33	0,30
orchard (hectares)	0,1	0,07	0,04	0,05	0,11	0,28	1,13	0,11
cattle (heads)	0,00	0,00	1,00	1,37	2,38	5,76	25,97	5,05
pig (heads)	1,00	1,51	2,24	4,03	7,79	14,12	46,44	5,01
horse (heads)	1,00	1,25	1,51	1,75	1,82	2,12	3,53	1,88
sheep (heads)	2,09	3,39	4,92	8,68	16,30	52,02	272,23	39,03
poultry (heads)	6,72	10,43	18,37	28,86	32,08	43,46	1700,20	43,38

References

Hungarian Statistical Review 2000. 78. volume
The Agricultural Census 2000. (CD-ROM)
Hungarian Agriculture 1851-2000 (CD-ROM)
Hungarian Agriculture, 2000. - regional data - (paper and CD-ROM)
Land Use of Hungary, 2000. - data by settlements - (paper and CD-ROM)
Livestock in Hungary, 31 March 2000. - data by settlements (paper and CD-ROM)
Agricultural fixed assets, 2000 – data by counties (paper)
Under preparation
Agricultural employment, 2000 – regional data (paper and CD-ROM)

The Agricultural Censuses of Egypt

Mahmoud Nazif
Agricultural Census Consultant for the Ministry of Agriculture, Egypt
FAO Consultant in Agricultural Statistics
e-mail: mahmoudnazif@hotmail.com

Abstract: Agriculture in Egypt constitutes a basic cornerstone of the socioeconomic structure. It still accounts for about 20 percent of GDP and absorbs about 36 percent of employment. Egypt has, so far, conducted six national agricultural censuses. The seventh agricultural census is now being conducted. This paper describes the sequence of activities for conducting the Agricultural Census 2000, namely: Census legislation – census committees - work plan – budget – organization of agricultural census – designing questionnaires and forms – preparing instruction and definition manuals - tabulation plan – pre-test surveys – pilot census – organization of field work – training – publicity campaign. – Census fieldwork - processing, tabulation and dissemination.

Keywords: Agricultural Census

1. Introduction

Egypt, officially the Arab Republic of Egypt, is a country located in northeastern Africa and southwestern Asia. It is bounded on the north by the Mediterranean Sea, on the east by Palestine, Israel and the Red Sea, on the south by Sudan, and on the west by Libya. It has a total area around one million sq. Km.
The vegetation of Egypt is confined largely to the Nile Delta, the Nile Valley, and oases.
The population of Egypt (the 2000 mid-year estimate) was around 66 million.

1.1 Agriculture

Agriculture in Egypt constitutes a basic cornerstone of the Socioeconomic Structure. It is the largest contributor to the overall development and progress of the society. Its role in Egypt is of great importance, as it provides livelihood for more than half of the population engaged in production, marketing and processing.
Agriculture still accounts for about 20 percent of GDP and absorbs about 36 percent of employment. Moreover, agriculture possesses many positive characteristics and potentialities.
The cultivating area yields as many as three crops a year, giving Egypt abundant agricultural yields.

1.2 Egyptian Administrative Divisions

The country is divided into 26 Governorates, 200 districts and about 5000 villages and cities. 13 Governorates are located in Lower Egypt, 8 in Upper Egypt and 5 in the Desert Region.

1.3 Agricultural Censuses Historical Background

The first agricultural census was conducted in Egypt according to the international agreement signed in Geneva on 14th of December 1928 between Egypt and International Institute of agriculture in Rome. According to this agreement the participating countries should carry out the census regularly every ten years. In 1945, The FAO became the successor of IIA.

Egypt has, so far, conducted six national agricultural censuses by complete enumeration of all agricultural holdings in 1929, 1939, 1950, 1961, 1982 and 1990. The seventh one, census 2000 is being conducted now. The Ministry of Agriculture (MOA) has the responsibility of conducting all these agricultural censuses.

2. The Agricultural Census 2000

2.1 Its role

The agricultural census is a part of integrated system of agricultural statistics with objective of providing data on structure of agricultural sector as well as socioeconomic information.

The basic objectives of census are:

- To Provide aggregate tables for fundamental agricultural data to use as benchmark for inter-censal estimates.
- To provide a frame of administrative units, buildings, households, and holders for other agricultural sample surveys.
- To provide data for small administrative units
- To obtain detailed data on livestock characteristics.

The agricultural census provides data satisfying the needs of government agencies, which could be utilized in planning and monitoring. This may contribute to regional planning and to optimum distribution of resources, as well as the needs of private sector.

2.2 Coverage

The agricultural census 2000 is conducted in complete enumeration covering all agricultural holdings in the country, with or without land, in rural, urban and desert areas. The census involved about 9 million buildings, 3.8 million agricultural holders with land and a total area of 8.8 million feddans or about 3.7 million ha. (1 feddan = 0.42 ha.).

2.3 Time Reference

The agricultural census is conducted for the agricultural year 1999/2000. The agricultural year starts on the first of November and ends on 30 October the next year.

2.4 The sequence of taking the Census

2.4.1 Census legislation

Statistics and censuses legislation has been issued in 1960, and has been emended by law 28/1982. It urges people to collaborate with the census staff and assures them of confidentiality of any information.

Penalties are imposed mainly for two kinds: in the case of respondents, for non-participation or for giving false information; and for violation of confidentiality on the part of census staff.

2.4.2 Census Committees

An agricultural census steering committee has been established. It includes various national ministries and agencies as well as directors of census organization and census users.
The committee studied the work plan and approved it. A technical sub-committee was established to provide technical advice concerning definitions, concepts, methodology, questionnaire designing ... etc. Also, a publicity sub-committee has been established.

2.4.3 Work Plan

The work plan involved agricultural census phases and basic activities in a time frame with appropriate and adequate time for each operation.
The workload of field operations was measured using the number of buildings and the number of agricultural holders from previous censuses and pilot census.

2.4.4 Budget

It was broken down in the following categories:

1. Personnel:
2. Production supplies and materials
3. Equipment.

2.4.5 Organization of Agricultural Census:

The general administration of agricultural census was organized in the following structure.

- Technical and training administration.
- Field survey and tabulation administration.
- Editing, analysis and dissemination administration.

2.4.6 Questionnaires and forms

A. Map

A map of a village or a city serves to identify the location of residential areas and natural barriers. Moreover, maps are used to determine the area of work (called segment) of each enumerator.

B. Basic questionnaire (the frame)

This includes the following items:

- Location of holding.
- Number of buildings and blocks.
- Names of the household's head and holder.
- Total area of the holding.
- Number of cows, buffaloes, camels, sheep, and goats, poultry and agricultural mechanical machinery.

C. Detailed questionnaire:

It includes the following items:

- Holder's name, his address, age and sex.
- Educational status and main other occupation of holder and number of members of the holder's household.
- Location of holding, number of parcels, land tenure and land utilization.
- Winter, summer and "Nile " crops and vegetables by kind and area.
- Fruit trees by kind, area and number.
- Timber, aromatic and medicinal trees by kind, area and number.
- Interplant crops and vegetables by kind and area.
- Number of wells and springs.
- Cultivated area by main irrigation source, and by main type of irrigation.
- Type of drainage in cultivated land.
- Number of cows and buffaloes by sex, age and purpose of breeding.
- Number of sheep and goats by sex and age.
- Number of poultry in house breeding.
- Chicken for meat production and hens for egg production.
- Hatching laboratories.
- Specialized farms classified under "other poultry".
- Other animals.
- Number of beehives.
- Agricultural mechanical machinery by kind and source.
- Number of agricultural workers in the holding during the agricultural year.

D. Other census forms

A number of other forms were used during the census operation in order to record aggregated data by segments (enumerator's area of work), villages or cities, districts, governorates and republic. It also used to report supervisor’s remarks.

2.4.7 Instruction and definition manuals

Two manuals were prepared, one for the first phase and the other for the second phase of the census. Each manual contained:

- Definitions of terms used, description of questionnaires and forms, and instruction for their completion;
- Methodology of field work;
- Rights and responsibilities for each level of the census field personnel.
- Census legislation and decrees.

2.4.8 Tabulation plan

The tabulation plan has been developed to identify the information which has high priority, and which should be released as early as possible to be most useful to data users. Other results which could be released successively.

2.4.9 Pre-Test Surveys

Pre-test surveys were carried out in the field on a small scale so as to test the suitability of questionnaires in actual field conditions.

2.4.10 Pilot Census

A pilot census was taken in three districts during the 1997-1998 agricultural year. One district was selected in Lower Egypt, a second in Upper Egypt and a third in the Desert Region.

2.4.11 Organization of FieldWork

Approximately 21,000 persons participated in executing the agricultural census. The work was organized according to the following four levels:
See flow chart page 10

2.4.12 Training of field personnel

Questionnaires and manuals were the basis for training. Field training was essential, and at the end of each training course the most qualified persons were selected through a written examination.

- Intensive courses were given to train top level staff.
- Other courses were provided to train Heads of Enumerators and enumerators at the governorate and district centers.

2.4.13 Publicity Campaign

Plans for publicity took place before and during fieldwork. The following media were used for the publicity campaign:

- Posters and pamphlets distributed to schools, mosques and educated farmers.
- Publicity through Friday speech in mosques and churches.
- Information in schools during the first lesson.
- Conferences, radio and television messages, plays, newspaper propaganda and editorials, agricultural extension campaign in villages and cities using loudspeakers in markets.

2.4.14 First Phase of the census fieldwork

The first phase of the census fieldwork, which started in December 1999, produced a complete list of buildings, households and holders by means of the following operations:

- Drawing maps: Each Head of Enumerators and his group working in a village delineated the boundary of the village and its components: buildings, hamlets, roads, canals, agricultural area, etc., using a sketch map form. This operation served also to allocate the segment to each enumerator.
- Each enumerator assigned serial numbers to all buildings in his segment.
- Listing of households and agricultural holdings: after numbering each building of a segment, the enumerator registered the name of the head of the households and wrote down the names of the agricultural holders and some basic data about the holding.
- Data editing: The above-mentioned frame was adjusted by using data from other sources.

2.4.15 Second phase of the census fieldwork
The second phase started at the end of the agricultural year with the objective of interviewing all agricultural holders using a detailed questionnaire, which included about 300 items. Data collected were edited in the field and in the office.

2.4.16 Processing, Tabulation and Dissemination
The first phase was tabulated manually to obtain results on the level of segments, villages and cities, districts, governorates and for the totality of Egypt. The second phase is tabulating mechanically but the number of basic items was tabuled manually too.

2.4.17 Final results
Final results will be printed for each governorate and tabulated by district. A volume will also be devoted to present aggregated data for Egypt. These books will be released successively and distributed to ministries, governorships, universities, institutes, libraries and other institutions.

3. Development of agricultural censuses of Egypt

3.1 The Agricultural Census 1929
It was the first census. In preparation for that census the agricultural census consultancy committee was established. Enumerators were recruited from Mayor assistants. Data was collected in 3 field phases:

The characteristics of the census:
1. Maps were not used and buildings were not numbered.
2. No social or identification data about the holder were collected.
3. A large number of questionnaires were used (20 questionnaires and forms)
4. The holding was considered all parcels operated by the holder within the boundary of the same village.
5. The desert Governorates were not included in the census.
6. The report comprised 16 tables; one of which was classified by size class of holding. All tables classified by district, governorates and Kingdom.
7. The number of agricultural holdings with land amounted to 1,213,915 and the area of holdings 7,420,696 Feddans or about 3.1 million Ha.

3.2 The Agricultural Census 1939
New items were introduced concerning quantities of fertilizers used by type of fertilizer and by season. Fruit trees were divided into fruitful and unfruitful. The number of permanent workers on holding from member of household and the number other than household members by age and sex was estabilished.

3.3 The Agricultural Census 1950
1. Desert Governorates were included.
2. Data on main occupation of holder, number of parcels and drainage were introduced.

3.4 The Agricultural Census 1961

1. Census and statistics legislation was established.
2. A pilot census was carried out.
3. Buildings were numbered.
4. Enumerators were recruited from masters and teachers of primary schools; they amounted to 5000 personnel
5. A new definition of holding was introduced. The holding was considered to be all parcels operated by the holder within the boundary of the same district.
6. A publicity campaign was better planned and made more intensive.
7. New items were introduced concerning age and legal status of holder, number of temporary workers, insecticides, amount of fuel used by hired agricultural machines, as well as appointing a specific date as time reference for livestock

3.5 The Agricultural Census 1982

1. The stages of census have been reduced to 2 visits instead of three.
2. Enumerators and supervisors were recruited from agricultural engineers.
3. Sketch maps for villages and cities were introduced.
4. Field staff books were introduced to summarize data collected and to report supervisor's remarks.
5. Data of the first phase was tabulated manually to be disseminated early on the level of villages and cities, districts, governorates and Republic.
6. The questionnaires and forms have been simplified and reduced from 20 to 8. The items on fertilizers, insecticides and fuel have been deleted. On the other hand, the items concerning total number of holder's member of household, level of education and sex of holder as well as inter planted crops have been introduced.
7. Written examinations at the end of the training program were introduced.
8. A separate report for each governorate was disseminated on the level of district and governorate. A general report was prepared on the national and governorate level.

3.6 The Agricultural Census 1990

1. The number. of districts for pilot census was increased to 3 instead of 2.
2. Training was made intensive; accommodation facilities were introduced.
3. Publicity campaigns were more effective.
4. Fisheries, fish cage and area under protective cover and poultry in houses were introduced.
5. It was the first census for which the Ministry of agriculture had the responsibility of processing and tabulating the data.
6. The report comprised 62 tables, 21 of which were classified by size class of holding.

3.7 The Agricultural Census 2000

1. The agricultural holdings were classified into holdings located in old land and holdings located in new land.
2. It is planned that the report will comprise more tables specially that classified by size class of holding.

3. The preliminary results show that the number of agricultural holdings in the census about 4.7 million, out of which 3.8 million holdings with land. The area of holdings is around 8.8 million feddans or about 3.7 million (ha).

Figure 1: ***Organization of The Census of Agriculture 2000***

Central Bureau
(Level of Republic)
Governorate Bureau
(Level of governorate and district)
Chienf of Enumerators
(Level of cities and villages)
Enumerators
(Level of enumeration areas)

References

FAO (1995).Programme for the World Census of Agriculture 2000.
Ministry of agriculture of Egypt (1933). The results of agricultural census 1929.
Ministry of agriculture of Egypt (1944). The results of agricultural census 1939.
Ministry of agriculture of Egypt (1957). The results of agricultural census 1950.
Ministry of agriculture of Egypt. The results of agricultural census 1961.
Ministry of agriculture of Egypt. The results of agricultural census 1982.
Ministry of agriculture of Egypt. The results of agricultural census 1990.
Ministry of agriculture of Egypt. The preliminary results of agricultural census 2000.
Nazif, M. (1971). Improving the questionnaires and methodology for data collection to enhance efficiency of the agricultural census. Ministry of agriculture of Egypt.
Nazif, M (1992). Development of agricultural census in Egypt since 1929 agricultural census. Ministry of agriculture of Egypt.

The National Agricultural Censuses in Developing Countries and the Programme for the World Census of Agriculture 2000

A.K. Srivastava
Indian Agricultural Statistics Research Institute, New Delhi, India,
Tel.: +91 11 5755508, e-mail: aks@iasri.delhi.nic.in

O.O. Ajayi
Federal Office of Statistics, Lagos, Nigeria
Tel.: +234 1 2670356, e-mail: ooajayi@infoweb.abs.net

Abstract: Experiences on Agricultural Census of some developing countries, with special reference to the FAO Programme for the World Census of Agriculture 2000 is the main theme of this paper.

Keywords: Agricultural Census, Developing Countries

1. Introduction

An agricultural census is one of the important activities in the agricultural statistics system of a country. The programmes of World Censuses of Agriculture have played a pivotal role not only in the conduct of Agricultural Censuses throughout the world but also in helping the countries in developing their own agricultural statistics systems. Maintaining an overall uniformity in the concepts, standards and definitions on the one hand and keeping the specific structural and logistic features of individual countries on the other, has been a challenging task in the conduct of agricultural censuses in different countries. Experiences of some of the developing countries in this context with special reference to programme for the World Census of Agriculture 2000 is the main theme of this paper.

2. Agricultural Census

The focus of the guidelines for the World Census of Agriculture (WCA) as given by the Food and Agriculture Organisation (FAO) of the United Nations (UN) was to measure the structure of the agricultural sector and at the minimum every ten years on a complete coverage basis. However, in many developing countries, coverage has been more or less on large sample basis and no more often than every ten years because of cost, nature of the agricultural practice and technical requirements of a total count. For the developing countries where subsistence agriculture is in practice giving rise to very many small holders, the census has always been a "sample census" as a practical strategy for covering the sector. This makes it possible in some of these cases to measure beyond the structure but additionally measure area and production of crops and the rural economy plus the demographic characteristics of the agricultural population. The institutional arrangements for the census has in practice included statistical bureaus, agricultural institutions and other stake holders. There were, of course, variants of these practices in the developing world.

3. Agricultural Censuses in Some countries of Asia

Most of the countries in Asia have been participating in programmes of Agricultural Censuses. Most countries conduct the censuses once every ten years. However, some countries such as Japan, Republic of Korea and India conduct it every five years. Most

Table 1 : *Agricultural Censuses in Some Asian countries*

Country	Times in past	Last Time	Next Scheduled	Survey Design		
				Complete	Partly	Sampling
Bangla Desh	4	1996	2002	#		
Bhutan	-		2000/01	#		
Cambodia	-					
China	1	1997	2007	#		
Fiji	3	1991	2001			
India	6	1995/96	2000/01		#	
Indonesia	4	1993	2003			#
Iran	4	1993		#		
Lao PDR	1	1999			#	
Myanmar	2	1993	2001			#
Nepal	4	1992	2002			#
Pakistan	3	1990	2000		#	
Philippines	8	1991	2002		#	
Sri Lanka	5	1982	2002		#	
Thailand	4	1993	2003		#	
Vietnam	1	1994		#		
Japan	10	1995	2000	#		
Korea	5	1995	2000/01	#		

countries participated in the WCA 1990 and are planning for a 2000 round. China has recently conducted its first Agricultural Census. Bangladesh, Bhutan, China, Iran, Vietnam have adopted complete enumeration while India, Lao PDR, Pakistan, Philippines, Sri Lanka and Thailand have followed a combination of complete enumeration and sample enumeration approach. Indonesia, Myanamar and Nepal have followed the approach of sample enumeration. Table 1 presents a brief description of conduct of Agricultural Censuses in some Asian countries (Ikeda, 2000). Survey items on which information were collected included holdings, areas, labour force, harvested areas production/yield, fruit trees by varieties, livestock by species, machinery facilities, land tenure, irrigated areas, etc. However, not all of these items were covered by every country. There might be other information sources for some of these items. It may be seen that the information included in these surveys are broadly in conformity with the ten categories of census items as suggested in WCA 2000 document.

4. Agricultural Censuses - The Indian Experience

India has got an advantage of having a sound system of land records including village maps which has formed the base for Agricultural Statistical System in the country. Agricultural Census is a major component of the system of Indian Agricultural Statistics. The agricultural operational holding is the ultimate statistical unit and it is through the agricultural census that basic data on important aspects of agricultural economy for all the nearly 100 million operational holdings are collected. So far six such censuses have been conducted ,the last one being for the reference year 1995-96, while the first one was conducted in 1970-71 Since the second agricultural census 1976-77, an Input Survey is also being conducted as a part of the agricultural census to provide State level estimates of consumption/application of various inputs by different size classes of holdings. The reference year for such surveys is one year after the reference year of the census. Thus the reference year for Input Survey for the sixth agricultural census 1995-96 has been 1996-97.

In India the main advantage of the data generated by these censuses has been the availability of information according to the different size classes of holdings, thus helping in policy formulation particularly for the small and marginal farmers who constitute the overwhelming majority of the country's farming community. The census has also been helpful in providing the basis for the development of a comprehensive integrated national system of agricultural statistics with links between its various components of the national statistics system.

Since the issues relating to Indian agricultural censuses are dependent on the system of agricultural statistics prevailing in the country as well as on the need for developing data base system, in an integrated manner, we discuss this system first.

The collection of agricultural statistics is a by-product of the normal departmental activity of the State Governments which are primarily responsible for the compilation of these statistics. Within the States, the collection of such statistics is primarily by the staff of the Department of Land Records. This organisation is not uniform in all the States. In fact, the States can be broadly divided into two categories viz. (I) temporarily settled and (ii) permanently settled. In the former, land revenue is fixed for a definite period of years and is subject to revision at the end of this period, while in the latter, the rate has been fixed for perpetuity. An elaborate revenue agency extending to the remotest village exists in the temporarily settled States which has already been referred to as land record States. Normally, in a group of 3-4 villages depending on their size, there is primarily a reporter which is known as Patwari or Lekhpal or Karnam or by some other names depending on the States concerned. This person resides locally and is well acquainted with the agricultural and economic conditions in the area under his jurisdiction. There are Revenue Inspectors, each incharge of the circle and above them are the supervising officers at the level of Tehsils/Sub-Division and Districts which form principal administrative units of the States. It is this organisation which is given the responsibility of the collection and compilation of agricultural statistics. In the permanently settled States, already referred to as non-land record States, there is no such elaborate financial agency excepting of the officers at Thanas/Sub-Division, each officer being incharge of 100-150 villages.

In most of the land record States, as far as acreage statistics is concerned, the area has been cadastrally surveyed. In these States, acreage statistics are available by complete enumeration and are obtained with a fairly high degree of accuracy. In the areas which are cadastrally surveyed but do not possess primary reporting agency, acreage statistics are available only by sample surveys. There are also around 7% areas which are un-surveyed and the estimates available are based on personal knowledge of the officers concerned.

A Timely Reporting Scheme (TRS) was introduced for timely reporting of estimates of area and production of principal crops, so as to reduce the time lag and delay in sending the crop estimates by the State Governments. Under this Scheme, the villages in each stratum are divided into five independent non overlapping groups each comprising 1/5 of the total number of villages. In the land record States where acreage statistics are collected by complete enumeration, one set of randomly selected villages is chosen for crop inspection on priority basis immediately after the sowing in each season has been completed, but in advance of the period prescribed in the land records manual for such crop inspection. It is this 20% of the villages which are covered under the agricultural census. However, in States where TRS is not in operation, the villages are selected independently.

The system of conducting agricultural census in India was initiated as a part of the 1950 World Agricultural Census (WAC) for which the data required were collected through sample surveys carried out by the Directorate of National Sample Survey which gave estimates for the country as a whole and for the States. These estimates, however, were of limited value for local level planning.

The first agricultural census with the agriculture year 1970-71 (July to June) as the reference period, was conducted as a part of the 1970 World Agricultural Census Programme sponsored by the FAO. It was based on the concepts and definitions introduced by the FAO as well as on complete enumeration basis by retabulating the data already available in the land records, in so far as land record States are considered. In the non-land record States of West Bengal, Orissa and Kerala where no such revenue agencies exists, the data were collected through well designed sample surveys.

The second agricultural census was planned to be conducted in 1975-76 but had to be postponed to 1976-77 due to administrative reasons. In this census, while the information on the number and area of operational holding was collected on complete enumeration basis, for the rest of the items, the information was collected on sampling basis. An Input Survey in a sample of 2% of villages was also carried out for the first time, as a part of this census. In this survey, data relating to the use of various inputs such as fertilizers, manures, pesticides, livestock and agricultural equipment and machinery were collected.

The third and fourth agricultural censuses were conducted with reference period as 1980-81 and 1985-86 respectively. As in the past, these censuses were also conducted on complete enumeration/sampling basis for the number and area of operational holdings. For other characteristics, the information was collected on sampling basis.
The fifth and sixth agricultural censuses were conducted with reference year 1990-91 and 1995-96 respectively. The methodology of these censuses were broadly along the lines of the previous census.
One of the main features of the item coverage in the Indian Agricultural Census is that it is a combination of a complete count and a sampling approach with emphasis on complete coverage on only two items. In fact, in states having a comprehensive system of land records, only number and area of operational holdings are obtained through complete count by retabulation of information already available in the land records. Other characteristics like tenancy particulars, land utilisation, sourcewise irrigated area under crops irrigated and unirrigated, are collected on sampling basis in 20% of the villages. In other states not maintaining land records, information from 20 to 25 per cent sample of villages are collected by household enquiry methods.
As already mentioned earlier, since the second agricultural census 1976-77, an Input Survey has also been conducted as a part of the agricultural census. The primary objective of this survey is to provide levels of consumption of various inputs viz. Fertilizers, pesticides, farmyard manures, agricultural implements and machinery, livestock and agricultural credit by five size classes of holdings i.e. marginal, small semi-medium, medium and large.
The sampling design for such surveys is normally a stratified two-stage random sampling with block/tehsil as the stratum, village as the first stage and the operational holding as the second stage of sampling. The number of villages covered is 7 per cent of the total number of villages in the State. These 7 per cent villages are selected randomly with tehsil/CD block as the stratum from the 20 per cent villages already selected in the main census. Four operational holdings are selected randomly from each of the five size classes of holdings. The data are collected through field enquiries from the selected operational holders. The survey covered the whole country and all types of agricultural holdings except institutional holdings. The reference period is one year later than the main census years and from July to June next year. The data are collected in two visits separately for Kharif and Rabi seasons immediately after the agricultural operations are over, in order to minimize informant bias on account of memory lapse. For Rabi season, the enumerators would cover the same operational holdings which were covered in the Kharif season.
For livestock, India has got a quinquinnial livestock census.

5. Experiences of Methodology Used in Agricultural Censuses particularly in Africa and in other developing countries

The survey methodology being adopted in carrying out the census of Agriculture in Africa in particular took cognizance of the coverage and peculiarities of the attributes of the population. The census had therefore been conducted using a very large sample rather than complete enumeration and was thus styled "sample census". The conditions that exist in Africa are probably similar to what the rest of developing countries experience in terms of coverage and scope. The frames of holders have never been available and in Africa they are generally small scale subsistence farmers. Quite often they have to be covered through large samples.

In Lesotho, the Agricultural census has always been through large samples. Basically it is a two-stage design, self-weighting with the first-stage being Primary Sampling Units (PSU's) which are combinations of Enumeration Areas (EA's) as demarcated for the Population Census exercise, and the second-stage units being households. Holdings within selected households were all studied. EA's have been combined into PSU's so as to achieve a first-stage unit large enough for various sub-samples for survey modules and over the 3 years of the usage of the master sample. For the most recent agricultural sample census, 110 rural PSU's and 20 Urban PSU's were covered with 27 households in the second-stage. For annual surveys, these numbers came down to 80 rural PSU's only. Stratification is by districts and ecological zones and PSU's proportionally allocated to the 24 strata so formed. When sampled PSU's are classified into Districts, Urban PSU's distribution outcome was one PSU per District except in Leribe and Maseru where 3 and 9 PSU's have been respectively selected. Apparently the sample has not been designed to estimate for the Urban component of the survey items.

In Mozambique, the design is similar to that of Lesotho; a two-stage design with Primary Sampling Unit (PSU) as the first stage and households as the second-stage unit. In their last exercise, the total sample covered 22,000 households. The office was generally being supported by donor agencies like the World Bank. In earlier censuses, there was no complete geographical coverage. In the 1993 census when the war just ended, 20 districts out of 129 were covered. The second round of this was followed up in 1994 and 30 districts with some overlaps with the 1993 coverage were covered. In 1996, 61 districts were covered studying 3904 households. The experience here was that the coverage was on sample basis and the whole country geographically could not be covered at the same time due to a variety of reasons.

In Uganda, the Ministry of Agriculture conducted an Agricultural Censuses with assistance from FAO of UN and again they were conducted on a sample basis. Not much had happened several years in the most recent past as the country depended on unsystematic routine data collection. In a new plan, the Ministry would be cooperating with the new Uganda Bureau of Statistics (UBOS). This country's experience typified the situation in Africa when the Census could not even be conducted because of lack of resources and when they were conducted, it took place with external resources and assistance.

Botswana represented a different approach. During the last population census of 1991 it had added basic agricultural items on the Population Census Questionnaire. This had enabled the country to obtain an insight into the structure of agriculture. The same had been planned for the 2001 population census and this exercise would continue to be supplemented by annual surveys. The annual surveys are canvassed through a sample which also is a two-stage design.

Egypt's effort was close to the practice in the developed countries where Agricultural Census consisted of a complete enumeration. The last census of 1990 used a frame made up of towns, villages and Agricultural co-operatives and all holders were covered. Items like areas, and production of crops, irrigation and drainage, employment, farm machinery plus livestock, poultry and beehives were covered.

6. The Nigerian Experience

Nigerian experiences at Agricultural Censuses go as far back as 1950, when it participated in the 1950 World Census of Agriculture. For lack of adequate finances, technical and administrative resources it was not possible to carry out a complete census, so a sample survey was conducted. Nigeria also participated in the 1960 WCA by conducting a series of sample surveys spread over a period of five years (1955 - 1960). The surveys covered northern Nigeria (1955-58), western Nigeria (1958-59) and eastern Nigeria (1959-60). Because the surveys were conducted in different times and at different regions, adjustments were made to estimates obtained earlier than 1960 to bring the estimates to the 1960 level. In 1962, a plan for an expanded programme of current agricultural statistics was developed with the assistance of the USAID. Under this programme, regular annual surveys had started and this was when the entire country was covered for the first time on sample basis simultaneously using objective measurement methods. From 1955/56 to 1963/67, the scope consisted of the following items, namely: Acreage and production of various crops and economic trees, rural market prices, number of livestock by sex and land utilization. From 1963/64, the scope was expanded to include rural household consumption enquiry for food and capital goods, prices, received and paid by farming households, land tenure, farm equipment, transportation, fertilizers used and demographic information. The 1974/75 sample census was designed to collect information on the structure and output and input of agriculture in the country. All land holdings in the country were covered partly on sample basis and partly by complete enumeration of the specialized and large holdings. 1984/85 Sample Census: This was Nigeria's participation in the programme of the WCA-1980 and also a necessary follow-up of Nigeria's attempt of the sample census carried out with national coverage in 1974/75. One major objective was to measure the structure and output of agriculture in Nigeria which could then be updated by the regular annual Rural Agricultural Sample Surveys (RASS). The scope like the 1974/75 census included Household composition i.e. age, sex, educational level attained, estimated annual income etc of household members, types of crops grown (area and production), land tenure pattern, type of livestock kept, type of poultry birds kept, quantity and value of fish caught, use and sources of fertilizers, manure and improved seedling and reasons for non-use. Under the sample design each state of the country was stratified into an average of 10 domains of study with one or a combination of two Local Government Areas (LGA'S) being defined as a domain. A two-stage design was then adopted for coverage of the traditional holders with the first stage unit being Enumeration Areas (EA's) and the second-stage unit being households. Apart from this, the design has been multi-phased, taking advantage of the evidence of greater variability in area measurement than in yield measurement. The sample at the first-stage was therefore broken into 3 phases, namely:-

1st Phase-Household listing and crop identification; covering all initial sample EA's and detailed interview of farm practice on a master sample selected from all the EA's.

2nd Phase-Area measurement on a sub-sample of EA's (1 in 8 of 1st phase sample EA's)

3rd Phase-Yield measurement on a sub-sample of EA's (1 in 5 of second phase sample EA's)

The EA's in the first phase were selected with equal probability and in each EA a constant probability of selection of households was determined to give a sample of about 10 households for detailed study. The 1993/94 Sample Census had three principal objectives, namely:

1. To provide structural data on agriculture in Nigeria mostly on those aspects that do not change frequently. In the context of this census, agriculture has been defined to include crop production (temporary and permanent), livestock rearing, keeping of poultry and fishing.

2. To obtain the socio-economic characteristics of the household, health and educational status, detailed demographic information and housing status. It was also to provide Local Government Area baseline statistics.

3. To obtain production figures at the state and national levels.

The census was in two phases; the first was to meet the first two objectives of the census while the second phase was to meet the third objective of the census. Also the census had two components, namely; the survey of modern holdings of agriculture which is covered on a census basis and the multi-phase sample survey of traditional small-scale holdings. Administrative data source was used to supplement the survey data particularly in respect of forestry.

SURVEY DESIGN: Domains of reporting - It was decided that the census should provide estimates for each LGA. There were 589 LGA's in the then 30 states and the Federal Capital Territory (FCT). States and FCT were domains of reporting at the next higher level and of course the national estimates reported for the whole of the country. Since planning in Nigeria is generally at the LGA and higher levels of administration, the census data met most of the planning needs.

The sample design called for a 2 phase sampling procedure with a 2-stage sample in each phase. In both phases the 1st stage sampling unit, generally called the Primary Sampling Units (PSU's) was the Census Enumeration Area (EA) and the second-stage unit was the housing unit (within which all households, were covered) and within selected households, the survey covered every agricultural holding. The EA's selected for phase II were a sub-sample of those of phase I. In the phase II EA's, the holders selected were the same as those selected for phase I.

There was also the decision that the General Household Survey (GHS), as far as possible, shared the same design as the National Agricultural Sample Census (NASC), not only for economy in data collection but also in order to allow for cross-analysis of the two surveys. This decision influenced the NASC sample design as regards, the 2nd stage sampling in Phase II.

The sample design was basically a 2 - stage sample design with the second phase sample selected in 10 replicates of 20 EA's/replicate. To allow for rotation of the sample, 6 replicates were studied for each year through the lifetime (5 years) of the master sample.

In all of the experiences of sample censuses of Agriculture in Nigeria, since 1950, there are still unsolved sampling problems in Agricultural Surveys. They are: how to treat distant farms, holders with dual residence, problem of seasonal migration, the coverage of Nomadic Livestock, the measurement of yield of tuber crops and boundary crops. The length of this paper will not allow for exhaustive treatment of those problems. Looking at the African region, Egypt's methodology of the agricultural census was similar to the experiences of some Asian countries, particularly the Chinese first census which was through a complete enumeration. Botswana obtained the structure of agriculture as a component of the population census. For most countries in Africa, the census has always been through a large sample. There is probably something to learn from Africa by the Asian countries in particular China. A sample census will profile the structure of agriculture reliably plus the added reason of economy. The Nigerian example clearly recommends itself. Also the basic agricultural items could be attached to the population census questionnaire. Botswana recommends itself in this approach. The coverage of socio-economic situation in both the Nigerian and Chinese examples is an issue that WCA guideline could consider.

There is evidence that African conditions are reasonably close to the Asian situation (China, India) and therefore might wish to recommend a combination of African approaches. Average size of holding in Africa was 1.6 hectares which was the same in China while India had 1.7 hectares. Household members per holding were in the same range in Africa and China, 5.8 and 4.1 respectively. Many small scale farmers (holders) existed with no list for them (no frame) sample coverage for them could have been satisfactory for China like in Africa. Or two types of questionnaires (the long and short) could be developed for China with the short questionnaire administered on a census basis while the long questionnaire is applied on a well designed sample.

Going by the World Census of Agriculture 2000, many African countries went beyond the structure of agriculture and extended their coverage to production. Some even conducted socio-economic surveys along with the census. But of course, nearly all the WCA in Africa were on sample basis. Many African censuses covered the sub-sectors of fisheries and forestry to a limited extent but the WCA guidelines have not demanded these. For most samples studied in the African Censuses, households were often selected and holdings within the households studied. This was the practical way of reaching the holdings which the WCA-2000 recommended for sampling. There is the additional benefit of analyzing the data based on households.

7. World Census of Agriculture 2000- some observations

Agricultural censuses must be viewed as an important component in the Agricultural Statistics System of the country. The World Censuses of Agriculture provide a basis for uniformity in the concepts, standards and definitions relating to Agricultural Censuses. Under these broad guidelines the individual countries adopt the censuses according to their own needs as well as resources. The WCA documents have been relevant to the requirements of countries as well as covering emerging areas of interest. Documents relating to inclusion of employment, aquaculture etc in WCA programmes are relatively recent additions. In WCA 2000 information on the role of women in agriculture as well as the inclusion of environmental and sustainable development issues are relatively current areas of interest. There is a growing realization that information on important agri-environmental indicators is essential to understand the environmental aspects of sustainable agriculture. It is also a fact that in most of the countries, environmental

statistics is obtained from secondary sources with other primary interests. The inclusion of environmental aspects in agricultural census is, therefore, a welcome proposition.
At this juncture, when technological changes are influencing every national and sub-national activity, its role need not be emphasized in the conduct of Agricultural Censuses. Applications of Remote sensing and Geographical Information System as well as the role of Information Technology has to be identified and brought out more clearly in conduct, analysis as well as dissemination of Agriculture Census results. Despite various constraints, even developing countries are organising their efforts to keep pace with developments in these technological developments. Food and Agriculture Organization has a pivotal role to play in sensitizing the issues relating to Agricultural Censuses in this context.
The FAO might wish to examine some special areas of Agricultural Census and give further guidelines for conduct of WCA, namely:

1. The role of women in Agriculture through Agricultural Census;
2. Production figures to be obtained through farmer's interviews;
3. Improvement of questionnaire content to include social and economic conditions of the agricultural population as Nigeria and China did in their census;
4. Need to conduct WCA 5 - yearly;
5. Attention to some special problems like Nomadic Livestock coverage, Distant farms coverage etc. in the guidelines.

References

Ajayi, O. O. (1995) *Design and Organization of the 1993/94 National* Agricultural Sample Census (NASC) of Nigeria.

Ajayi, O. O. (2000) *Survey Methodology for the Sample Census of Agriculture in Nigeria with some comparisons of Experiences in other Countries* paper presented at the International Seminar on China Agricultural Census Results, Beijing.

FAO (1995) *Programme for the World Census of Agriculture 2000,* FAO Statistical Development, Series 5.

FAO (1996) *Conducting Agricultural Census and Surveys*, FAO Statistical evelopment, Series 6.

FAO (1997) *Guidelines on employment*, FAO Statistical Development, Series 5a.

FAO (1997) *Guidelines on the collection of structural aquaculture Statistics*, FAO Statistical Development Series 5b

FAO, Accra (2000) Mission Report to Lesotho Bureau of Statistics

FOS (1989) FOS Agricultural Statistics Development, Sources, Scope, Nature and Prospects

FOS (1992) National Agricultural Sample Census Design

Ikeda R (2000) *Survey Methodologies in Selected countries of the Region*, Lecture delivered in FAO/SIAP Workshop on Multiple frame Agricultural Surveys held at Bangkok from 22-29 November, 2000.

(1998) *All India Report on Agricultural Census 1990-91,* Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India, New Delhi.

(2000) *All India Report on Input Survey 1991-92,* Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India, New Delhi.

Larson, L.O. and Narain, P. (1998) *International Standards And Definitions for Agricultural Statistics and Requirements for World Census of Agriculture*, Paper presented in International Conference on Agricultural Statistics 2000, held at Washington D.C., 18-20 March 1998.

Panse, V.G. (1966) *Some Problems of Agricultural Census Taking with special reference to Developing Countries*, Food and Agricultural Organization of the United Nations.

UNECA (2000) Mission Reports on the Evaluation of Addis Ababa Plan of Action for Statistical Development in Africa.

6 June 2001

Invited Paper Session

Data Sources and Design Issues for an Agriculture Statistics System

Chair: J. Gallego

The New Economics of Remote Sensing for Agricultural Statistics in the United States

George Hanuschak, Robert Hale, Michael Craig, Rick Mueller
National Agricultural Statistics Service, U.S. Department of Agriculture

Galen Hart
President, Galen Hart Associates, Virginia, USA

Abstract: The National Agricultural Statistics Service started an applied research program in the early 1970's to utilize space-borne satellite imagery and digital data, such as Landsat, to identify crop type and measure crop acreage for selected States in the U.S. In the initial research, software development for information extraction was the key activity. NASS needed the traditional remote sensing outputs of cover type and probability of being correct. In addition, NASS needed an unbiased (or a negligibly biased) statistical estimator of crop area at the State and county level. Using Purdue University's LARSYS (early 1970's version) system as the base, NASS and the University of Illinois's Center for Advanced Computation developed a customized software package called EDITOR to accomplish these tasks. NASS staff converted the software across several platforms and called it PEDITOR. PEDITOR gives conventional remote sensing categorization (or classification) outputs and a double sampling regression estimator as well. Over the years, NASS staff have continually improved PEDITOR by adding functionality and efficiencies.
Most recently, NASS staff have emphasized system efficiency with expert system like features to ease the labor burden of analysts. In addition, recent improvements in the county level estimates have been added as well. The system now runs on high end Windows NT desktop computers. The program is currently used in seven States plus a minority student outreach program with Florida A&M University. In the last several years, NASS has developed a Cropland Data Layer (CDL) in geographic information system (GIS) format (See Figure 1) for public use. It is being used by GIS proficient users for watershed monitoring, agribusiness planning, prairie water pothole monitoring, crop rotation pattern analysis and animal habitat monitoring.
The CDL can be viewed and ordered on CD at
http://www.nass.usda.gov/research/Cropland/SARS1a.htm

Keywords: remote sensing, agricultural statistics, cost, economics, partnerships

1. The Old Economics (1970's-1980's)

The economics has changed rather substantially over the last three decades. The initial statistics and software research in the early 1970's was dominated by research staff costs and computer processing costs. The first full State project, Illinois in 1975, cost $750,000 but still included substantial research and statistical and software development costs. The first applications costs were established for processing the State of Iowa (1978) in time for end of season crop area estimation. At that time, the total cost per State was estimated at $300,000 and the data source was Landsat Multi-Spectral Scanner (MSS) at 60-80 meter resolution, with 4 spectral bands of data.

By the early 1980's, there were eight States being processed. The total cost per State was $150,000. Major item costs were associated with remote sensing analyst staff (although spread out over 8 States which increased efficiency substantially over one or two States with about the same staffing levels) and the cost of processing the data on mini-computers and on the ILLIAC and Cray supercomputers. Data costs were $300 per Landsat scene and generally a small portion of the total project cost. When Landsat thematic mapper became available in 1982 at 30 meter resolution with 7 spectral bands of data, the program faced a fork (applications continued or a research program and switch to Landsat TM) in the road and only had resources to go in one chosen direction. At first, NASS continued the eight State application using Landsat MSS data, with a small amount of resources devoted to research on Landsat TM. Even with the limited research program resources, it became apparent that the accuracy levels could be improved substantially by using Landsat TM but at a higher cost for the data and for the information extraction (more bands and better resolution). Thus in 1988 due to budget cuts and U.S. government plans to not have MSS data on future Landsat or other earth resource observing satellites, NASS abandoned the eight State application using Landsat MSS but increased the research program with Landsat TM.

A new applications program was initiated in 1991 in Arkansas and Mississippi. In 1992 Louisiana was added to the program. Budget cuts hit again and Louisiana and Mississippi were dropped from the program. In fact in 1995, the program was only intact for Craighead county in Arkansas. Needless to say is that the program had a very rocky road at that point in time. Thus, it was apparent that NASS had to partner and resource share with other Federal and State government Agencies in order to continue and expand the program again and add a public cropland data layer for distribution as well as the internal statistical products. In preparation for this, NASS analyst staff emphasized a more automated and more expert system like version of PEDITOR to ease the labor burden on analysts. When this was adequately accomplished, NASS was in a position to seek out Federal, State and University partnerships.

2. Resource Leveraging Partnerships

Starting in 1997, a new and key data license partnership between NASS and USDA's Foreign Agricultural Service (FAS) and the Farm Services Agency (FSA) was established. Landsat 5 data could be used for projects that serviced these agencies programs. Thus, the Landsat data costs for a program expansion were reduced. In addition, with Landsat 7, data costs at EROS Data Center were set at $600 for a system geo-referenced data set without licensing restrictions. USDA FAS has a USDA-wide license for Landsat 7 with Radarsat for $405 per scene. Thus, the Landsat data costs were no longer a major constraint as when the cost for Landsat 4&5

peaked for government buyers at $4,400 per Landsat scene in the early 1990's.
The highest cost historically was for expert remote sensing and statistical analysts to run the complex PEDITOR package for both remote sensing outputs and for statistical estimates with relative sampling error outputs as well. NASS had a small group of expert analysts in its' Research Division who were doing centralized analysis for several States. However, it was recognized that a decentralized analysis staff directly in NASS State Statistical offices would expand the Agency's analytical capabilities. In addition the analyst will have the advantage of more localized knowledge of the crops and cropping practices and other sources of data to evaluate the relative contribution of the Landsat to the NASS crop area estimation program at the State and county level.
A local analyst (with a stable position and longevity) and State office management will be in a better position to service other State government and university and farm organizations and agribusinesses by providing them a Cropland Data layer in geographic information system (GIS) format. For example, the analyst in Illinois did both Illinois and Indiana in a regional concept which is most cost efficient. Small two to four State regions will likely be the most cost efficient model.
This data layer has already proven valuable to data users of NASS who often combine the NASS cropland data layer with other data layers. Some examples are for water quality assessments for watersheds, and location plans for a new agribusiness facility, such as a soybean crushing plant and for grain storage and transportation planning and for prairie water pothole monitoring in North Dakota**.**
Since 1998, NASS has added the States of Illinois, Indiana, Mississippi, New Mexico, and Iowa and continued to do North Dakota and Arkansas as well. Florida A&M University and NASS also entered into a cooperative program for minority student outreach in remote sensing and GIS by analyzing a portion of northern Florida. Establishing and maintaining effective and win-win partnerships for the Cropland Data Layer is challenging though. The analyst position needs to be strongly supported by all the partnering organizations and needs some longevity (5 years or so) to be effective. Some of the partnerships are stronger and more stable in terms of longevity than others. A companion paper by Craig (ASPRS 2001) entitled "A Resource Sharing Approach to Crop Identification and Estimation" goes into more detail about each partnership for the Cropland Data Layer program.

3. The New Economics (late 1990's and 2000)

NASS has entered into a new economics for crop area estimation and a Cropland Data Layer for public consumption through the new resource leveraging partnerships. The total cost per State is now $75,000 and should drop further with the addition of more States. See the attached Figures 2 and 3 for the graphic representation of the cost reduction over time (inflation adjusted and non-inflation adjusted). Major reductions were realized in the cost of the Landsat data and even more dramatic reductions occurred in the cost of data processing which has gone from mini-computers and supercomputers to high end PC desktops now. All the processes, including full Landsat scene multi-temporal pixel categorization, are done on PC's The non-inflation adjusted total project cost has been cut in half (See Figure 2). When adjusting for inflation (See Figure 3), the cost reduction is more than fivefold from the first large scale application in the early 1980's. The current program and the early 1980's are the only comparable periods to compare costs over. The reason is that the number of States that labor costs are being spread over are very similar in number and scope of effort. Dramatic cost reductions have been achieved in the cost of the Landsat data and for the data processing of the Landsat data to categorize it into most likely crop types (See Figure 4). Spreading out the labor costs over as many States as possible and having co-funding partners is the key to future cost reductions.

Perhaps more importantly, there are now more accurate and valuable output products, both internal to NASS and external to NASS data users. These include a State level estimate with relative sampling error reductions on the order of two to five compared to the ground sample data only. County level (small area) estimates with measurable sampling errors have recently been improved through several alternative estimators. Third, there is a new publicly available Cropland Data Layer at the 30 meter pixel level categorized by crop type and formatted for standard geographic information systems input.

The use of the Cropland Data Layer by GIS proficient data users outside of NASS is a major new thrust. In North Dakota, the NASS State Statistical Office (SSO), the North Dakota State University and an agri-business firm used the Cropland Data Layer as one input into a key agri-business decision. The location of a soybean processing plant was being evaluated. By combining the Cropland Data Layer with a transportation network layer and other layers such as commercial grain storage facilities for soybeans, a decision model was built by the University and used in the agri-business decision for the plant location. Another anticipated major use of the Cropland Data Layer is in watershed monitoring and in fact this is one of the major reasons for the North Dakota State government to show interest in the program. The State of Illinois chose to add to the Cropland Data Layer by adding other ground data on land covers and converting it to a Land Cover Layer.

4. Private Sector Perspective

The economics of private sector remote sensing for agriculture and renewable natural resource monitoring has followed a somewhat similar pattern to the USDA/NASS experience with a peak in the 1980's followed by a downturn and then a substantial recovery at the turn of the century. In the late 1970's and early 1980's, a number of commercial applications were developed. Many of these applications dealt with clustering and classification of Landsat Multi-Spectral Scanner (MSS) data (60 meter to 80 meter resolution) into land cover types, including agricultural land covers in some cases. Many of these applications produced large scale land cover maps for government and private sector use. However, a number of these products had limited accuracy assessments and were often restricted by the lack of a statistical sample of ground data for verification purposes. Usually, a small amount of ground data or photo-interpreted aerial photography was used for verification and creation of the classification matrix (omission and commission errors). The attempted commercialization of Landsat, in the commission early 1990's, came next and sharply increased prices per Landsat scene (over

$4,000). A number of the firms with Landsat applications such as large scale land cover were forced to scale back or, in some cases, go out of business. Several firms diversified further into value added products that went beyond classification and mapping. Some were as simple as enhanced raw or classified image products put on glossy color prints. Other value added products were more complex, such as using geographic information systems to combine Landsat data with other data layers to customize products/solutions for clients.

Among the better known remote sensing, geospatial information firms in the Unites States are Earth Resource Satellite Corporation and Pacific Meridian (now a unit of Space Imaging Corporation). These geospatial information companies, along with many others now, often "mix" multiple data sources in GIS format to customize products/solutions for their clients. For example, one could combine weather data contours (precipitation, temperature, wind etc.), soils data contours, satellite vegetative index data and crop condition data to get a crop yield forecast.

Some of the data sources are satellite image data, aerial scanner digital data, U.S. Geological Survey digital maps, digital terrain maps, digital soils maps, official government statistics and ground data with global positioning systems (GPS) locations etc.

With the launch of Landsat 7 in 1998, a new era began with government subsidized data pricing at $600 per scene, system registered and license free. This was a big boost to all Landsat applications, and both public sector, private sector and universities benefitted from the reduced prices. Large scale applications took off again as large area studies were no longer severely restricted by data prices. At the recent Landsat Data Continuity Mission Workshop, held at the U.S. Geological Survey and co-sponsored by the National Aeronautics and Space Administration (NASA), a number of data distributors,users and value added firms reported new growth opportunities created by Landsat 7. One specific value added firm example presented at the conference stood out, as Metapath Software International (MCI) reported an annual revenue increase of 277% over the previous year. MCI is a value added firm that does worldwide urban analysis with Landsat 7 (pan-sharpened and license free) as one of the major ingredients in their geospatial information extraction process. Many of those attending the Workshop expressed a desire for Landsat 8 to mimic the total success of Landsat 7, that is high quality data, reasonable pricing, similar spectral coverage, near nadir views, fixed polar orbit and wide swath coverage.

Agribusiness, particularly the large international commodity firms, are using remote sensing as one input for crop forecasts and estimates in statistically "under served" production areas of the world. Their interests range from their own forecasts of large area produced crops (grains, fiber and oilseeds) to more limited area crops (palm oil, cacao, tea and coffee). Their expenditures for this type of information development are largely unknown due to the proprietary nature of their business. However, those who have experience in the area agree that hundreds of thousands or more dollars are spent annually by these firms for their geospatial information extraction, including crop forecasting and estimation.

Similar to the described USDA/NASS experience, more frequent temporal coverage would benefit private sector geospatial information extractions firms with agricultural clients. Data coverage is a major issue in timely crop monitoring applications. In many rain fed crop areas of the world, better than 16 day coverage is required. In some tropical areas, radar data is being explored because of the lack of optical sensor data coverage due to excessive cloud cover during fixed satellite overpasses. Three potential future data providers (Resource 21, Matra Marconi and RAPIDEYE) have identified agriculture as a primary commercial market. If any of these ventures succeed and have reasonable pricing structures, then that should be another boost for agricultural applications of space-borne remotely sensed data.

5. Satellite Data Sources

NASS staff have a very strong preference for the Landsat data. The nadir look, the wide area image swaths, and the spatial resolution and spectral bands are virtually ideal for the NASS Cropland Data Layer and crop estimation program. The only issue with Landsat is the temporal frequency. NASS staff desires eight day coverage as a minimum and would prefer even better temporal coverage. Backup to the Landsat for the NASS program is the Indian IRS satellite series, which have many Landsat-like features. SPOT has not proven to be a cost effective backup for the NASS program. NASS has used SPOT for area sampling frame construction, on occasion, where price reductions are available through State consortiums or now under a USDA-wide license or existing credits from USDA/FAS.

NASS gets questions about its plans for using very high resolution satellite data. At this time there are no major plans to use very high resolution data, such as IKONOS in the NASS program other than for occasional small quality control checks. NASS has farmer reported data from a stratified area frame sample to evaluate the Landsat data. The swath width of IKONOS is too small for wall-to-wall coverage needed for the Cropland Data Layer and the cost per square mile too high relative to Landsat. This only covers NASS needs as the IKONOS satellite provides very high quality data for many site specific applications, such as mapping a city.

6. Future Perspective (2001-2010)

NASS will continue to pursue partnerships, primarily with State governments, to expand the crop area estimation and Cropland Data Layer program to more States. The desires are for program expansion to the top 15-20 total cropland States in the United States. The Midwest, the Delta and the Great Plains are likely areas for expansion as resources allow for. Expansion beyond that point is unlikely.

The value of the Cropland Data Layer to the general public is hard to quantify but is considered quite substantial to those proficient GIS data users who combine it with other data layers to solve their problems of interest. Examples of uses to date are watershed monitoring, prairie water pothole monitoring in the Dakotas, grain transportation and storage planning, animal habitat monitoring, agri-business plant location analysis, farm equipment dealer planning and crop rotation pattern analysis.

Acknowledgements

The authors would like to extend their thanks and appreciation to all the analysts in the States involved. Those are Patrick Willis in Illinois, Dath Mita in North Dakota, Ken Akins in Mississippi, and Dave Christiansen in New Mexico. In addition, the State Statisticians and their staffs for extensive ground data preparations for the States in the program. Thanks also go to Martin Ozga for recent improvements in PEDITOR aimed at reducing analyst labor, including a new mosaicking algorithm. Thanks to Jim Mergerson, Charles Day and several members of the Area Frame Section for assisting in analysis preparations or analysis.
The authors would also like to extend their appreciation to Carol House, Fred Vogel, Joe Reilly, Rich Allen and Ron Bosecker of NASS for their headquarters executive management support of this program and to Allen Vandergriff of USDA/FAS and Glenn Bethel of USDA/FSA.

References

Allen, J. D., and Hanuschak, G. (1988). The Remote Sensing Applications Program of the National Agricultural Statistics Service: 1980-1987. SRB Staff Report Number SRB-88-08 USDA, NASS, Washington, DC-USA

Bellow, M. (1994b). Large Domain Satellite Based Estimators of Crop Planted Area. In: *1994 Proceedings of the Section on Survey Research Methods*, August, 1994, American Statistical Association, Toronto, Canada, pp.194-199.

Craig, M. (2001). A Resource Sharing Approach to Crop Identification and Estimation. *2001 ASPRS/ACSM Annual Convention Technical Papers, available on the CD, Proceedings of the ASPRS 2001 Conference,* Bethesda, MD-USA.

Day, C. (2001). A Compilation of PEDITOR Estimation Formulas. RDD Research Paper RDD-01-01, USDA, National Agricultural Statistics Service, Washington, DC-USA.

Fishman, J. (2001). The Landsat 7 Applications of Metapath Software International (MCI), USGS/NASA Landsat Data Continuity Workshop, Reston VA-USA.

Graham, M. (1993). State Level Crop Area Estimation Using Satellite Data in a Regression Estimator. In: *Survey Methods for Businesses, Farms, and Institutions*, ICES Part I, NASS Research Report No. SRB-93-10, USDA, Washington, DC-USA, pp. 7-11.

Hanuschak, G., Allen, R., and Wigton, W. (1982). Integration of Landsat Data Into The Crop Estimation Program of USDA's Statistical Reporting Service. *Proceedings of the 1982 Machine Processing of Remotely Sensed Data Symposium,* Eighth International Symposium, LARS, Purdue University, West Lafayette, IN-USA, pp. 45-56..

Mueller, R. (2000). Categorized Mosaicked Imagery from the National Agricultural Statistics Service Crop Acreage Estimation Program. Available on the CD: *Proceedings of the ASPRS 2000 Conference*, ASPRS, Bethesda, MD-USA.

Ozga, M., and Craig, M. (1995). PEDITOR - Statistical Image Analysis for Agriculture. Presentation at the Washington Statistical Society (WSS) Seminar, April, 1995, USDA, National Agricultural Statistics Service, Washington, DC-USA.

Ozga, M. (2000). Batch Processing of Remote Sensing Jobs on the PC. Available on the CD: *Proceedings of the ASPRS 2000 Conference,* ASPRS, Bethesda, MD-USA

Vogel, F., and Bange, G. (1999). Understanding USDA Crop Forecasts. In: NASS/WAOB Miscellaneous Publication No. 1554, U.S. Department of Agriculture, Washington DC-USA.

Figure 1: *Cropland Data Layer*
McLean County, Illinois

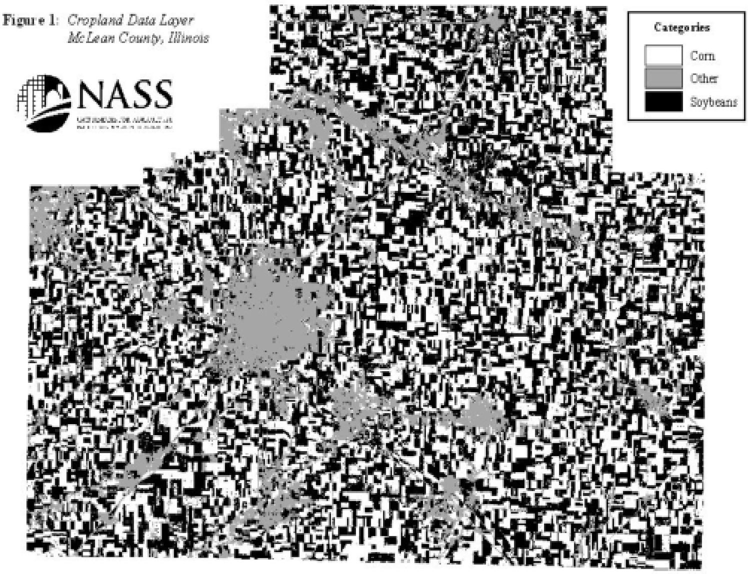


Figure 2: *Cost Per State (Unadjusted)*

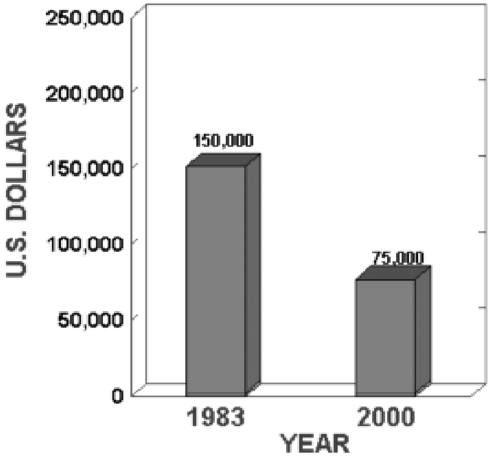
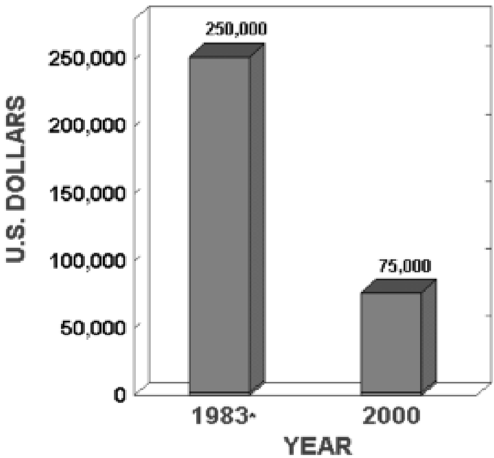
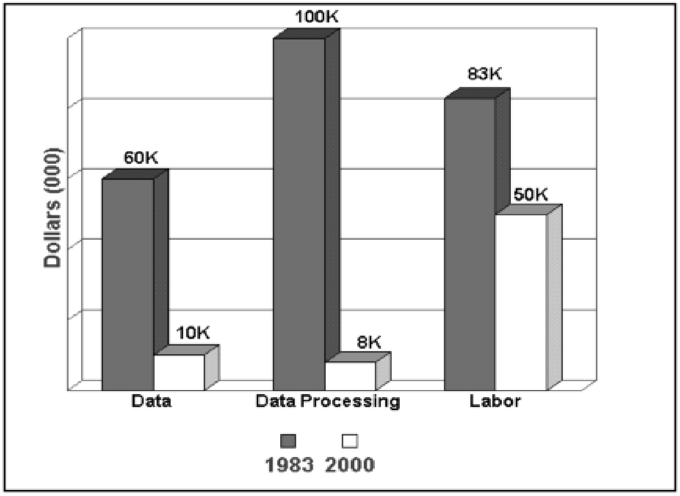


Figure 3: *Cost Per State (Adjusted for Inflation)*



*1983 adjusted to year 2000 dollars

Figure 4: *Major Item Costs (Adjusted for Inflation)*



The EU Farm Accountancy Data Network (FADN): Partner in Statistics

Krijn J. Poppe, Karel van Bommel
Agricultural Economics Research Institute (LEI), The Hague
e-mail: k.h.m.vanbommel@lei.wag-ur.nl

Abstract: FADNs and Statistics are natural partners that benefit from each other. This partnership can be widened and deepened to mutual benefit. FADNs have been used for many years as a tool for monitoring and analysing agricultural policy. To perform this task they gather micro-data and depend on statistics to make them representative. In recent years, with shrinking budgets for agricultural statistics and new information needs in the agricultural domain, FADNs have also become a supplier of statistics. FADNs could benefit from statistical methodologies like post-stratification and panel techniques. Quality control is also an issue. This paper presents also a number of cases where FADN and statistics are involved. In addition, current developments in the EU-FADN are described. This can help statisticians to influence and improve the partnership.

Keywords: FADN, farm accountancy, statistical sources

1. Introduction

This paper discusses farm accountancy data networks (FADN) as a partner in agricultural statistics, based on European experience. FADNs have been used for many years as a tool for monitoring and analysing agricultural policy. To perform this task they gather micro-data and depend on statistics to make them representative. In recent years, with shrinking budgets for agricultural statistics and new information needs in the agricultural domain, FADNs have also become a supplier of statistics, hence the description 'partner in agricultural statistics'.

After a short introduction on FADNs in this section, we discuss current developments in the FADNs and then discuss statistical aspects. This is illustrated in section 5 with some cases and followed up with a discussion from the viewpoint of statistics on quality and efficiency.

In some EU members, FADNs have existed for 50 years and more. Public concerns on food, be it with quantity or quality, and -more recently- agricultural production methods and rural areas as well as -in the past- the importance of the farm vote, have resulted in a relatively high level of government intervention in the agricultural sector. This policy making implies a need for policy relevant research and data. In addition to statistics, the FADNs played an important role to satisfy these needs. The detailed methods of government intervention in modern times (cross compliance, modulation) makes this more true than ever: there is a huge demand for micro economic data that can be used for simulations. An FADN has the advantage over normal statistics that it is a panel that can answer questions with multiple variables like 'what is the effect of a lower payment

per dairy cow for farms larger than 50 cows and a stocking rate above 2 on the farmer's income, his environmental performance (e.g. mineral balance) and the EU budget'.
Data for FADNs are gathered by local accountants, supported by instructions at national and (for the European Union) on EU level to reach harmonisation of results. Individual, but anonymous, data is transferred to national organisations and to the European Commission (DG-Agri) in Brussels. As the European FADN was created on top of already existing national FADNs (France, Greece, Spain, Portugal and several candidate countries had to build up an FADN when they prepared for membership of the EU), the output is harmonised, not the process.

Table 1: *Two different types of FADN*

Aspect	Type X 'low cost - low value'	Type Y 'high risk - high value'
Central organization FADN	Ministry of Agriculture	Research Institute
Type of finance	internal budget	output-related
Data gathered by	sold by accounting offices	own staff
Farmer's participation	is paid	free
Feedback to farmers	low	high
Interest by farmers	low	high
Data flow and its:		
- information content	low	high
- innovation	low	high
Data used by research	incidently	often, critical success factor
Political culture	data monopoly by ministry no open access by others	policy advise; consensus building in public domain
Main role of EDI	can solve lack of interest	can reduce higher costs
Typical example	Germany	The Netherlands

National history and circumstances are reflected in different systems of data collection. In some countries research institutes (including universities) play an important role in data collection, in others the national ministry of agriculture buys (re-calculated) tax accounting data from accounting offices. As these systems differ in use, (marginal) cost structure and innovation behaviour, we can classify them with the help of two arch-types (Poppe et al., 1997). Table 1 provides a description of these two arch types. More information on the European FADN can be found on the web site of the EU (EU, 2001).

2. Current developments FADN

In this section we discuss some developments in policy making and accounting that influence the future of the FADN.

Enlargement EU

In 1993 the Copenhagen European Council made the historic promise that 'the countries in Central and Eastern Europe that so desire shall become members of the Union'. Eleven Central and Eastern European countries (CEEC) and three others have become candidate countries (CC). The EU's main instrument of assistance in preparation for adoption of the *acquis communautaire* is Phare. Several candidate countries use this program to build or adapt an FADN. The Czech republic and Hungary have probably made most progress. They have an FADN up and running but face still problems in linking the FADN with the FSS. The next group consists of countries that already have a tradition in collecting micro-data or have made progress in starting an FADN. This group includes Slovenia, Poland, Lithuania and Estonia. A third group has still to start. Due to diplomatic/bureaucratic reasons the guidance from DG-Agri to the CC has been less than the co-operation between Eurostat and the CC's statistical offices. This gap has partly been closed by consultancy from different member states (leading to the export of different systems from EU member states).

CAP reform and Farm Return 2000+

As the CAP reforms, new types of data, e.g. on environmental performance, food safety and off-farm income, are needed. There is also a need of information on the change of the role and functions of agriculture: rural business, including tourism, the link with the environment and nature, forestry and other (new) activities on the farms. This means that the number of objectives and scope of FADN are increasing. At the same time it is remarked that FADN can not fulfil all these objectives. But if policy makers are interested in the relationships within the farm between e.g. policy measures (subsidies, quota, intervention prices) on one hand and the effect on income and (e.g. environmental) performance/behaviour on the other hand, the FADN is the most attractive method for gathering this data.

In 'Ricastings - the feasibility of a new farm return for the FADN' (Abitabile, 1999) a new farm return has been proposed to meet these needs. The Farm Return 2000+ should use up to date ICT-options to support conversion and speed up data exchange. There will be a distinction in the Farm Return 2000+ between:

- hard core of (obligatory) data (table 2);
- voluntary data (table 3).

Table 2: *Hard core data in Farm Return 2000+*

Survey	Topics
Farm structure statement	ESU, UAA, AWU, crop areas
Farm profit and loss account	Inputs, outputs
Farm subsidy statement	Subsidies per regulation
Farm flow of funds statement	Cash flow, investments
Farm balance sheets	Capital, liabilities

The voluntary data gives the opportunity for exchange between member states of data that is nationally available, without obliging all countries to gather those data. It is stressed in this context that also for the voluntary data a minimum quality level is required, for instance on the representativeness of the samples.

Table 3: *Proposal in Farm Return 2000+ for voluntary data*

Survey	Topics
Mineral balances	Nitrate and others
Cost of production	Gross margin, Physical data, Cost of production
Diversification on the farm	Organic production, Processing on the farm, Agri-tourism Landscape maintenance, Forestry
Activities outside the farm	Non-farm activities, Non-farm income/capital

Related to the distinction between obligatory and optional data, it is suggested to define different incomes:

- income from primary agricultural production;
- income from all activities on the farm including non or semi-agricultural activities;
- total income of the household, including off-farm income.

The Farm Return 2000+ should be based on accounting statements (balance sheet, profit and loss account, enterprise margins, etc.) which are familiar to accountants. This is also the form in which the data are published, but the collected data should be at a more detailed level than those published at the moment. This provides member states with an incentive to use harmonised indicators in their own publications; secondly, the member states and regional accounting offices can calculate these indicators and publish them directly when they finish their own accounting year. This speeds up the availability.
National farm returns differ largely between member states and it is not feasible to oblige member states to harmonise them in the short term. There is no common European tradition in farm accounting, and national tax systems influence methodologies greatly. There is also the difference between type X and Y in collecting the data. Nevertheless the EU should support the adoption of common farm accounting methods. It is most likely that (future) member states with not much tradition in farm accounting and not much interference with tax accounting will adopt such common methodology quicker and swing the balance in favour of using common methods.
The EU has officially adopted IASC standards to improve harmonisation. In this respect the release of an international accounting standard for agriculture has a good timing.

IAS 41
The International Accounting Standard Committee approved in December 2000 for the first time ever a new standard for agriculture: IAS 41. This standard prescribes the accounting treatment, financial statement presentation and disclosures related to agricultural activities. Biological assets should be measured at their fair value. Biological assets are defined as 'living animals and plants that are controlled by an

enterprise as result of past event'. The IAS makes a difference between fixed biological assets (e.g. milk cows) and current biological assets (e.g. fattening pigs). Agricultural produce that has been harvested, is no longer included in the biological assets, but as a product and the fair value should be determined at point of harvest. There is a presumption that fair value can be measured reliably for a biological asset. The IAS 41 dictates how fair value can be determined when there is no market price available.
A first test in the Dutch FADN (Boone et al., 2001) showed that not all biological assets have reliable market-prices. For example, there is no market for half-grown slaughter pigs or products on the spot market (lettuce). To use fair value, it is essential that there are market prices available. Even in well-developed counties this may become more difficult in the future, with more co-operation in the chain and higher diversification of products. Notwithstanding these drawbacks, for most products a reliable fair value can be determined. When all FADNs would use IAS41, the comparability will improve.

3. Current developments in ICT

An FADN is a data intensive operation. The breath-taking developments in Information and Communication Technology (ICT) have a big influence on the efficiency and quality of FADNs, and much work is to be done to exploit the opportunities of these developments.
The revolution in ICT is based on two developments: the price of storage capacity has fallen dramatically and computer processing power has increased enormous (and thus also becoming very cheap). The cheap storage capacity makes discussions on information needs sometimes ridiculous. Instead of making a choice on data to make available from an already existing administrative system, one could just as easy leave the decision to the end-user. The same is true for discussion on aggregations: in accounting many elementary data (e.g. one sale of wheat in October) are aggregated into yearly data. This is a loss of information value (due to aggregation and later availability) that is not necessary anymore. Recording the elementary data is enough and users can compute aggregated data when they need this, taking advantage of the cheap computing power.
Based on these trends in ICT the FADN (and accounting offices) in the Netherlands have improved their efficiency and quality by using electronic data from banks, auctions and herd books. By recording the business partner with whom the farmer conducted his sales and input-transactions, data can be provided on the links in the agricultural chain. In some countries (like Austria) an exchange of data with the IACS system (in which subsidies for farmers are registered) are tested.

4. FADN and statistics

Eurostat publishes information on the structure of agriculture in its Farm Structure Survey (FSS), that are collected by the national statistical offices (Eurostat, 2001). EU legislation states that 99% of the standard gross margin (SGM) should be covered by the national FSS. As the structure of farms differs between member states, the threshold in Economic Size Unit (ESU) can differ. The Netherlands and Denmark have relative high thresholds.

The FSS is the universe of farms which is the basis for the FADN and helps to make it representative. The FADN has a field of observation, which represents at least 90% of the SGM of the country. This also results in different thresholds: from 2 ESU in Greece to 16 in the Netherlands (where 3 ESU is the threshold of the FSS). Every national liaison agency conducts his own selection. The selection plans are submitted to the FADN European Union's FADN management committee for approval.

Stratification
Within an FADN's field of observation, there is a great diversity of farming. Some farms are very large while others may be very small. Some farms concentrate on crop production, others specialise in livestock rearing while others practice mixed farming, that is, both crop and livestock production. On these two criteria alone, i.e. economic size and type of farming, the field of observation of EU farms is highly heterogeneous.
To ensure that the sample of farms adequately reflects this heterogeneity, liaison agencies stratify the field of observation before the sample of farms is selected. The reasons for stratification are the reduction of variance, efficiency and non-response-handling.
For FADN purposes the EU is divided into about 100 FADN regions. The European FADN makes also use of stratification and uses three criteria: region, economic size and type of farming. The number of farms in each cell is derived from the FSS organised by Eurostat. This survey employs the same typology as that used for FADN (EU, 1989).

Weighting system of EU-FADN results
Due to stratification, a weighting system is needed in the calculation of FADN results. It is based on the principle of 'free expansion': for each holding in the sample, an individual weight is calculated (extrapolating factor). In order to calculate this individual weight, holdings in the sample and in the field of survey are stratified according to the same three criteria. To produce indicators, weighted averages are calculated using these weights applied to each individual farm recorded in the sample.

Micro-lab
The liaison agencies have many micro(-economic) data available about the agriculture in the specific member state. The data is extremely privacy sensitive, as it is information about the farmer's income. In addition to national data protection acts, the EU's FADN legislation protects the participating farms. According to this, it is prohibited to use the information against the individual farm (e.g. by tax or environmental authorities).
This situation makes the access of individual data to potential users outside the FADN network problematic. France and the European Commission make queries on individual data for registered researchers possible, and use software to check that the output are averages based on a minimum number of farms. Following the example of Statistics Netherlands, the Dutch FADN starts this year with a so called micro-lab in which outside scientific researchers can get access under strict conditions. A mutual benefit for the FADN and university researchers could be breakthrough research and publications that the FADN could use in policy research.

Panel techniques
One of the fields for the statistical experts to explore is the possibility of panel techniques on FADN data. Panel techniques use a data-file that consists of information

about a number of observed items on a number of moments (e.g. yearly data for a period of 5 years). Use of FADN data is often based on the data of one year, or a time series from several years, without using the information that participating farms in the (rotating) panel provide several observations. This is uneconomic use of the data set. Panel techniques are especially useful with studies into small sample groups, for example organic dairy farms. The Dutch FADN contains about 25 such farms. This results in 25 observations for a cross section. When a panel approach is used for the years 1995 to 1999 the number of observations increases to 94.

Post-stratification
Farms are selected for the FADN on basis of the most recent, but often an outdated FSS. At the moment that FADN data are available, sometimes a newer FSS is available. Or the research is aimed at certain variables (e.g. stocking rates of cattle) that are not perfectly correlated with the original stratification variables, and on which additional information in an FSS (or other data set on the universe) is available. In these cases the quality of the research can be improved by post-stratification. Until now, this is not very much practised, perhaps because it asks for a very good link between FSS and FADN.

FADN = simulation tool for policy analysis
The FADN is first of all a simulation tool for policy analysis, with statistics as a by-product. But why not use this by-product for statistics as it is available anyway. Besides it is available, it can improve statistical publications with data that can not easily be collected by normal statistical surveys, like income-statements (privacy sensitive) or the use of pesticides. It is a costly operation to collect the use of pesticides from all farms, or a large sample without the data checking possibilities that the FADN has (Poppe, 1999). It is often better to collect detailed data only from a sample of farms (and use resources for checking and trust building with the farmer), than collect the data with less quality from as much farms as possible.

Standard results
The FADN standard results are a set of statistics that are periodically produced and published by the Commission. They describe in considerable detail the economic situation of farmers by different groups throughout the EU. The figures are published in a public database on the FADN homepage (EU, 2001).

5. Cases

In this section we provide some concrete examples of the trends we discussed in previous sections. We do not provide examples of policy research with the FADN, its traditional objective. Instead we concentrate on the relationship between the FADNs and statistics. As a by-product the FADN has always been in the business of statistics: the income data published by the FADN are of course statistical information. As these data are often more visible than the policy evaluation (that are often carried out internally at Ministries or DG-Agri and not published), it even leads to the wrong image that the FADN is a statistical operation. In recent years some statistical offices had quite some problems to explain why such a large part of the statistical budget should go to a small sector like agriculture. That is especially the case at member state level where the argument that data are needed for decision making in Brussels is not always convincing. Therefor some statistical offices (like in the Netherlands, Italy and Eurostat) have turned their eye to FADNs to provide input for their work.

Survey into low incomes
The EU's FADN has family farm income as one of its main indicators. As Hill (2000) correctly argues, this is not very informative on the real income situation of the family in comparison to other sectors. The Dutch FADN (and some others) also collect income on non-farm data and were recently asked to make a number of statistics on farm families 'in poverty' (although excluding their assets) to promote research in this area (Everdingen, Van et al., 1999). This would not has been easily from other sources, as e.g. Dutch tax data has only a limited number of agricultural families, without disclosing their type of farming and farm size; this would make it difficult to formulate (agricultural) policy recommendations.

FADN and price statistics
As agricultural chains make more use of contracts and product differentiation, spot markets become less representative. In cases where the FADN collects its data on a monthly or quarterly basis, it would be possible to use FADN data to judge the quality of traditional price statistics or even replace them.

Capital goods, environmental investments and national accounts
In the Netherlands the FADN provides Statistics Netherlands with several data for statistics. Data on pesticides and energy use are provided for statistics on these topics. The FADN concentrates on commercial farms, Statistics Netherlands on data gathering on very small farms (not represented in the FADN) and on additional questions. The FADN also provides data on types of assets (capital goods) and, also for the Dutch environmental agency, on environmental investments. The interpretation if an investment has an environmental character or not is not so straightforward, which makes a calculation on FADN data attractive above a questionnaire. FADN data are also used in national accounts.

Life Cycle Analyses
Life cycle analysis is a technique to calculate the environmental impact of a product from cradle to grave. One of the big problems to use it by food producers, is the lack of data at farm level, notwithstanding the large amount of agricultural statistics. Experiences in Switzerland and discussions in an international network (Weidema, 2000) suggest that FADNs can play a useful role in this domain.

FADN data for Eurostat's Agris
To support modelling and policy analysis, Eurostat is actively replacing its SPEL-database by Agris, a new database to improve the coherence between statistics without unauthorised estimations of missing data. In addition to the Economic Accounts for Agriculture there is a need to allocate inputs to activities (e.g. feed costs to keeping dairy cattle). One of the potential sources for this data is the FADN, for which an econometric model exists that allocates costs to activities, based on the individual data.

ARTIS: a new software tool for the Dutch FADN
The Dutch FADN had to renew its software over the last years and tried to take advantage of the developments in ICT, as described in section 3. A new data management system was developed (Poppe, 2000) that is characterised by flexibility and re-use of data in different contexts. Changes in a data model used to describe the data needed, can easily be mapped into changed data gathering procedures. This is used to collect data on a quarterly basis, where changes in data can easily be handled. Re-use of data is also supported: what in one research domain is called a tractor, can be a fixed asset in another domain (accounting data model), an energy-sink (energy data model) or a means of transportation (logistics data model) in other domains.

6. FADN and quality

Statistical offices stress quality; to accept statistics from the FADN, it should have a minimum level of quality. The most relevant norm for the definition of quality is the ISO 8402. This states that: 'Quality is the total of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs'.
The quality concept proposed for a new farm return is built on the recent developed quality concept for statistics from the European Statistical System and focuses on seven main components:
- Relevance (e.g. for policy making in DG Agri);
- Accuracy (e.g. guaranteed by stratification and accounting procedures);
- Timeliness and punctuality (e.g. correct data should be available in 9 months after a year);
- Accessibility and clarity of the information (e.g. internet site with meta-data);
- Comparability (e.g. standardisation between member states and in time);
- Coherence (e.g. with statistics like EAA, IAHS, FSS);
- Completeness (e.g. field of observation).

If statistical offices use FADN data to make and publish statistics, they will be interested in the quality of the data and the processes used to gather the data. Certifying FADNs could be an interesting option. For example, Statistics Finland grants institutes a label of quality, when its statistics are up to a specified quality level.

7. FADN and efficiency

The FADN is an expensive operation. But everything is relative: it might be expensive compared to agricultural statistics like price statistics or harvest estimations or even the FSS. Its costs however dwarf by the agricultural budget of the EU, that itself is the most important item in the EU budget.
The FADN provides users with a data set and a number of statistics that more traditional statistics like the Farm Structure Survey ever can provide. Some of the FADNs have a low marginal cost to provide extra data.
The FADN and the FSS are both conducted for the European Commission. The co-operation between the FADN and Eurostat has improved to a good level. But further deepening of the co-operation could be fruitful. We mentioned already the Agris project. The use of FADN in other statistics should be explored. Regionalisation of the

EAA, environmental statistics and quality control (coherence) come to mind. Harmonisation of definitions is than a pre-requisite and up to date ICT should be used.
FADN managers are often focussed on policy analysis and have a background in accounting, business economics or ICT. Statistics, with providing information on reliability (standard deviations) and meta data, are not the first priority, although much progress is made in several units in building web sites. Statistical offices could help in this, as well as in areas like post stratification and panel techniques.

8. Conclusion

This paper has argued and tried to show that FADNs are a valuable source for agricultural statistics. Although statistics are for the FADNs a by-product of the data set that supports ex-ante and ex-post policy analysis, it also provides interesting data for pure statistics. It needs statistics like the FSS to present representative results and it can learn from statistical techniques like analysis of panel data and (post-) stratification. More co-operation is therefor useful. The EU FADN could also co-operate with Eurostat in supporting candidate countries in adopting the relevant acquis communautaire.
FADNs can improve quality and efficiency by employing up-to-date with ICT. This can lead to more flexibility, re-use of data and new types of data. The work on the Farm Return 2000+ could therefore also be of interest for statisticians: to learn from it and to integrate it as a data source for statistics. Co-operation of the EU-FADN with non-EU ones (in Europe, but also in other OECD countries) could be beneficial for international (WTO) decision making.
In conclusion: FADNs and Statistics are natural partners that benefit from each other. This partnership can be widened and deepened to mutual benefit.

References

Abitabile C. et al. (1999) *The feasibility of a new farm return for the FADN*, LEI, The Hague.
Boone K. & Bommel, K. Van (2001) Fair value in agriculture - first implementation of IASC E65, in: *Pacioli-8*, K.J. Poppe et al., LEI, The Hague.
Eurostat (2001) Website: http://europa.eu.int/comm/eurostat
EU (1989) *FADN: an A to Z of methodology*, Brussels.
EU (2001) Website FADN: www.eu.int/comm/agriculture/rica
Everdingen W.H. van, Venema, G.S. & Bommel, K.H.M. Van (1999) *Agrarische gezinnen en hun inkomens*, LEI, The Hague.
Hill B. (2000) *Farm incomes, wealth and agricultural policy*, Aldershof.
Poppe K.J. et al. (1997) *Pacioli-4 Proposals for innovation of farm accountancy data networks*, The Hague.
Poppe K.J (1999) *The use of farm accountancy data networks in generating environmental data*, paper for the joint UN-ECE/Eurostat work session on methodological issues of environment statistics, Jerusalem.
Poppe K.J. (2000) Software for the Dutch FADN as a tool for micro-economic research, in: *Pacioli-7 Agenda 2000 and the FADN agenda*, K.J. Poppe et al., The Hague.

Weidema B.P. & Meeusen, M.J.G. (2000) *Agricultural data for life Cycle Assessments*, LEI, The Hague.

The Agricultural Statistical System in Italy[1]

Massimo Greco
ISTAT, Via Ravà 150 00142 Roma, Italy
e-mail: msmagrec@istat.it

Laura Martino
ISTAT, Via Ravà 150 00142 Roma, Italy
e-mail: martino@istat.it

Abstract: The system of agricultural statistics has recently been completely revised by the Italian National Statistical System. It has been conceived as a co-ordinate and integrated set of information on various aspects of interest deriving from the emerging trends in this sector. A complete methodological revision of the main surveys has been launched. It involved the setting up of a new sampling scheme with the farm structure survey sample playing the role of a mother-archive with all the other survey samples nested in it. The recourse to nested samples along with the set up of a computerized system for survey managing and the use of some non traditional statistical procedures (such as the use of area sampling frame) represent the most significant innovations introduced in this area.

Keywords: agriculture, statistical system, methodological revision

1. Introduction

Even though, in the last decades, agriculture has played a marginal role in the Italian economy giving little contribution in terms of the added value of the Country (3,4% in 1999 - ISTAT 2000), the radical economic and social transformations characterising the western Countries have recently caused a new and emerging interest in this sector. Such an interest is due not only to economic factors but also to some aspects related to the quality of life and the protection of the public health. In fact, the food industry suffers from phenomena that originate in the primary sector, such as the spreading of the ESB and of foot-and-mouth disease in the animal farms or the production of genetically modified organisms. The growing attention of the consumers to the quality of food products has strongly reinforced the need of looking at a single agro-food product as the result of a chain of processes linked together. In this approach agriculture doesn't represent just an economic sector but mainly the origin of the food chain and for this role it deserves special attention.

In this frame the figure of the farmer too has considerably changed over the years, tending towards a more managerial role compared to the past. Currently the agriculture activity is not only oriented to the land cultivation but is also integrated with other activities providing an important added value such as the agritourism, the direct sale of the farm products, and environmental protection, among others.

---

[1] Although the authors are both responsible for the whole paper, the paragraphs 4.1, 5.1, 5.2, 5.3, 5.4 and 5.5 have been written by L. Martino, all the others by M. Greco

Following these new trends, the Common Agriculture Policy has recently moved from a system based on production support to a new system whose leading concepts are production rationalization and income maintenance. Since the 44% of the total European Community's budget is still devoted to the accomplishment of the agriculture policies (40.993,9 millions of Euro in the year 2000 as compared to 93.280,4 millions of Euro in total - EUROSTAT, 2000), the Commission strongly needs updated, reliable and comparable statistical information in order to support the decisions adopted in the different sectors of competence. The request of statistical data in the agriculture field from the European Union is in accordance with the weight of this sector in the process of European integration. Among 222 legislative measures adopted from the European Community before 1997, 122 (55 %) regard the agricultural sector (ISTAT, 1998).
The renewed importance of agriculture at the economic and social level, related in Italy also with phenomena like the recovery and protection of the environment, has generated deep modifications in the information provided in this sector. Typical examples are some international projects, for instance the FADO (Food Agricultural Data Outline) project (Eurostat, 1999) and the TAPAS actions (Technical Action Plans for Agricultural Statistics) where the Member States are involved in defining new orientations in agriculture and informative needs coming out from them. The European Union is the most important but not the only international user of the statistical data. FAO, OECD, OMS, amongst others, are regular users of agriculture information. Moreover FAO, with its recommendations on the agriculture Census, has also a rule of proposal and of bond, for certain aspects, on the content of the surveys. A second important and parallel group demanding agriculture statistical information comes from the various public and private national users which use the survey results for the management of the own politics, economic and managerial activities or for studies and researches.
In conclusion the agricultural statistics are strongly conditioned by numerous, complex and largely diversified demands for information. For this reason recently Italy started to set up a new statistical system conforming to these requests of completeness, timeliness, efficiency and quality, based on a co-ordinate and integrate set of information on various aspects of interest (Schirinzi, 1998).

2. The reference points of the system

From the legislative point of view the first element of the system, at the general level, is the Order in Council of 6 September 1989, n. 322 regarding the reform of the statistics produced by public institutions in Italy. Through this order a national statistical system has been created with the aim of providing standardized official statistical information and carrying out the surveys and the collection of data included in the national statistical program. The law introduces two new important elements fundamental for the statistical production of our Country:

- definition of a set of official statistics with precise constraints for the agencies responsible to carry out the surveys (for instance respect of some quality standards, of common classifications and nomenclatures, rules on the confidentiality and the spreading of the results) and for the respondents (who are obliged to respond);

- identification of ISTAT as the center of the system but recognition of other actors as potential producers of official statistics (Regions, Ministries, Provinces, Municipalities, Chambers of Commerce among others).

At a more detailed level, the statement of the agriculture statistical system also takes further elements into consideration:

- the above-written FADO (Future Agricultural Data Outline) promoted by the European Union with the finality of promoting a debate on the directive lines of the evolution process of the actual European agriculture statistical system for the next 10-15 years and on the harmonization process of the objectives of the national systems.
- the TAPAS actions ((Technical Action Plans for Agricultural Statistics) settled by the Council of the European Union to improve the agricultural statistics, simplifying and rationalizing the existing production and developing new researches in this field;
- the agreement protocol ISTAT-MIPAF (Ministry for the Agriculture and Forestry Politics)-AGEA (Agency for funds' distribution in agriculture) to integrate, to harmonize and to rationalize the statistical information in Agriculture, Fishery and Forestry with the finality to answer at all the informative needs of international, communitary and national interest;
- the third agreement protocol ISTAT-MIPAF-Regions to find out the statistical needs, to define the activity to carry out and the relative modalities of execution and to settle the engagements to assume for each Institution;
- the new and increasing informative needs connected with transformations of the agriculture sector in Italy. These include:
  - a new approach in the production process and in the work organization, closer to that prevailing in other sectors;
  - strong links of the agriculture with activities close to the industrial production (transformation and distribution);
  - deep modifications in the social structure of the sector, reflecting on the traditional social and economic figures;
  - statement of the polyfunctionality of the farms;
  - the complex relation between farm and enterprise;
  - the spreading of extra-sectoral activities, integrated with the agricultural income;
  - the connection with advanced services;
  - the strong presence of the public expenditure as support to income;
  - the radical change of the institutional order and the reduction of the importance of the central national institutions (regional decentralization and European unification);
  - the change of the sectoral political objectives (control of the offer, quality, environmental limitations, rural development);
  - the need to join the formation and the monitoring of European politics.

3. Basic infrastructure of the agricultural statistical system

The representative model of the productive process of a modern statistical system includes three main elements: the sources (surveys, administrative data, etc.), the products and the services, and the treatment of the information. In this integrated model - constituted of input, throughput and output - many different sources and products are mediated from the informative system.

The scheme of the new Italian agricultural statistical system, as it has been designed, requires the constitution of a set of basic infrastructures to be used as a support to the surveys:

- the ASIA (Statistical Archive of Active Enterprises)-Agriculture archive;
- the land cover map;
- the area frame;
- the administrative archives.

The Asia archive includes the indicative data and some basic information on the farm (Filippini, 1998). It is updated through the information coming from surveys and by comparison with the existing archives of other Institutions. It represents the base to draw samples from in the period between two general Censuses.

The land cover map is the graphic representation on a map of the national territory that is divided into sub-areas according to a pre-defined land use classification (for instance wood, agrarian cultivation, fruit-trees, etc.). It allows obtaining information on the distribution of cultivation. Currently ISTAT is experiencing the use of a land cover and utilization map as a support for the stratification of the statistical units of a limited portion of the Italian territory based on a 43 items classification already adopted in the CORINE land cover project. In the same area, a pilot study has been carried out by ISTAT to build an area frame for the forest and agricultural sectors, following the experience of the LUCAS project.

The administrative archives of other Institutions are numerous and largely heterogeneous since they have been set up with objectives other than statistical production. Their utilization for statistical purposes is conditioned by the need to harmonize concepts, definitions and classifications. The utilization of administrative data would give important advantages in terms of money and time saving.

4. The surveys of the agricultural statistical system

4.1 Methodological revision

The new system of statistics on agriculture has not only involved a revision and an enlargement of the contents and of the number of the existing surveys but also their methodological revision (Benedetti, 2000). It has also required a deep methodological revision of their contents and of their processes to improve the availability and the timeless. The choice of the sample technique, wherever possible, explains this strategy. The main advantages of this choice compared to the census survey are the following:

- saving of human, material and financial resources;
- reduction of the statistical annoyance for the respondents;
- improvement of the quality of the information (better possibility of analyzing and checking the elementary data, reduction of the total missing data);
- time reduction in recording and processing data;

improvement in the timeless of the results diffusion.

The sampling frame used to carry out surveys in the agricultural sector has been totally revised in its composition and in its concepts and has been based on the use of a panel of farms. The aim was to avoid duplications in data collection and to allow complete integration and comparability of information derived from different surveys. The result

is the possibility to provide a complete representation of the farm features and their interaction and, from a statistical point of view, reducing the non-sampling error.
To match these targets, it has been decided to set up a survey system based on subsamples nested within a reference sample, the one used for the farm structure survey. Thus all the surveys on farms (e.g. economic results of the agricultural enterprises, animal livestock among others) provide information about statistical units already included in the farm structure inquiry. The main drawback of this scheme is that it produces some additional costs. They stem, on the one hand, from the need to manage the mother-sample as an archive to guarantee completeness and updating of the sampling frame and, on the other hand, from the need to maintain a strict timeliness of the mother-survey since the schedule of all the others depends on it
All the surveys use a panel technique with a yearly rotation scheme of 20%. This approach has many advantages. Firstly it gives the opportunity to follow the trend of the observed phenomena over time. Secondly it reduces the organization costs since most of the farms (80%) don't need to be approached every year. Thirdly it helps in getting more steady estimation and provides a useful tool to check the correctness of the elementary data by the mean of comparison of the same unit in different periods of time. Since surveys included in the agricultural sector are multipurpose, their optimal sample size and allocation of units to strata have been defined according to the Bethel method (1989). It represents a generalization of the classical optimal stratified allocation by Neyman (1934) but it allows introducing multiple sampling error constraints at regional and national levels in accordance with the European legislation.
Another important methodological innovation was the introduction of estimation techniques that allow the use of auxiliary information. The calibration estimator (Deville & Särndal, 1992) help in producing sampling estimates that are coherent with the ones obtained from other surveys or for the total population. Moreover it allows to reduce the sampling variance when highly reliable auxiliary variables strongly correlated with the variables of interest are selected. Basically it requires that some known totals have to be reproduced when sampling weights (direct weights multiplied by a corrective factor) are applied to the sampling units. To improve the efficiency of results it has been decided to use as auxiliary variables the values of the variables of interest at the preceding year.
Missing data have been integrated via sampling weights under the hypothesis of homogeneity of response within strata (Särndal, Swensson & Wretman 1992) or using calibration estimators.

4.2. The structural surveys

4.2.1 The farm structure survey

The survey on farm structure collects general information on the farms and data on cultivation, livestock, mechanical means, work used in the farm and external relations (ISTAT, 2000). ISTAT carries out these surveys annually although the European Union (Regulation 571/88) demands to execute it every two years. The objectives of the survey too are quite enlarged compared to the European Regulation. Such a Regulation requires to:

- collect, in the reference periods close at that of the farm structure (December), the conjunctural information listed in six European legislations and related to area cultivated with cereals, cultivation interesting the new CAP, pig production, cattle

production, sheep and goat production, and production and utilization of milk in the farm. These informative needs are satisfied through the inclusion of all the required questions in a single questionnaire. This unified approach allows a reduction in the number of the interviews and, consequently, the statistical annoyance for the respondents and to remove disagreeing data related to the same phenomenon;
- satisfy the demand of information on new subjects connected with the changes in the agricultural sector or study deeper specific aspects. This objective is realized through the adoption of a modular approach. Inside the basic questionnaire, indeed, one or more additional sheets can be inserted, changing year by year, collecting information on new or emerging fields. The advantages of this solution are the flexibility of the system and the rationalization of all the phases of the survey and of software and hardware resources.

Until today the additional module has dealt with the fruit trees (1997), the environment and the territory (1998) and social-rural aspects (1999).
Currently the survey includes around 80.000 farms. Such a sample size allows to provide reliable information over a large range of agricultural features.

4.2.2 The survey on the economic results of the farms

The adoption of the Regulation CEE n.2333/96 concerning the new European system of National Accounts (SEC95) has posed the issue of availability of economic information by single farm. To satisfy this demand, a survey on the economic results of the farms has been planned with the aim at obtaining data on the structure of the farm costs, on the investments and on all the other items of economic interest (contributions, extra-farm incomes, rents, interests, retributions, etc.)
The main features of the survey are:
- the annual periodicity;
- the integrability with the farm structure survey, due to use of a sub-sample of the main survey, that guarantees a better reliability of the results and a large amount of information on the same sampled units;
- representativeness at national and regional levels.

4.2.3 The surveys on the production means

The surveys on the production means concern the distribution of the mineral and organic-mineral fertilizers for agriculture use, of the pesticides, of the seeds and of the complete and complementary feeds.

4.2.4 Specific projects

One of the main characteristics of the Italian agricultural statistical system is the flexibility, that is, the capability to answer at new informative demands or to study deeper specific themes in relation with the evolution of the agriculture in time. In this context, new surveys and monographic researches have been carried out such as agritourism, flower cultivation, and quality production.

4.3 The conjunctural surveys

4.3.1 The surveys on the area and production of the cultivation

The informative needs relevant to the herbaceous and wood cultivation concern basically:

- estimations on the sowings and forecasts of the cultivation plants, both the ones subject to the CAP regime (cereals and other products) and the ones not subject to the CAP regime;
- estimations of production trends per cultivation during the agrarian year;
- definitive results on areas and production cultivation.

The methodologies used to satisfy these various needs are different. They always take into account the opportunity to economize the resources and to reduce the statistical annoyance for the statistical units. Currently ISTAT uses two different techniques: the estimation-based surveys and the sample-based surveys. In the last years, the possibility to use administrative data has also been explored. Another methodology, much more high-tech, used by the Ministry of the Agriculture and Forestry Politics within the AGRIT project (Consorzio Italiano per il Telerilevamento in Agricoltura, 2000), requires the use of a satellite. The estimation surveys provide provisional data on the trend of the area cultivated and of the productions in a short time at the provincial level. Definitive results are obtained at regional level of the area cultivated and of the productions of the vegetal products and of the main agrarian wood cultivation by the farm structure survey.

4.3.2 The surveys on the forest sectors

The surveys on this sector regard consistence of the forest resources, the interventions to preserve forest property and the prejudicial events as the fires and the infractions of the laws protecting the forest and the hydrogeologic environment.

4.3.3 The surveys on the animal sector

The surveys on this sector are characterized by a strong degree of integration and coherence with the farm structure survey, that - as already stressed before - represents the heart of the system. The inquiries are strongly limited by the communitary legislation which determines the main contents, the classifications and definitions, the periodicity of execution and some methodological aspects of the surveys. The surveys involve three sectors: animal livestock, meat production, and milk and dairy production.
The inquiries on animal livestock are entirely integrated with the reference survey of the system, the farm structure. In fact, data on cattle, buffalo, pig, sheep and goat livestock, provided in December as required by the communitary Directives 93/23, 93/24 and 93/25, are directly collected by the farm structure survey; while those on cattle, buffalo and pig livestock issued in June come from panel sample surveys obtained as sub-samples of that of the farm structure. Again the advantages of such an approach are the ones already listed for all the other surveys designed as a panel sub-sample of the farm structure sample.
Inside the reference scheme of the agricultural statistical system there are some surveys regarding the agro-industrial area strictly related at the primary sector. In detail, information on milk production and utilization in farm, demanded by the communitary Directive 96/16, are collected by the farm structure survey.
The inquiries on the slaughter-houses and on the dairies concern the first transformation of the animal products. Both of them have been recently completely revised (Martino & Greco, 2000). They are carried out monthly and use the sample methodology and CATI (Computer Assisted Telephone Interviewing) technique to collect data. The choice of the sample methodology depends mainly on the need to provide the European Union with the information requested in a short time (within 45 days for the milk and 60 days

for the slaughtering from the reference month). Considering the high number of the enterprises operating in these sectors in Italy, a total survey would not allow a full respect of the deadlines. Moreover the sample technique allows to better control quality of the information since a deeper check of the elementary data and a strong reduction in the total non-response has been made possible.

4.3.4 The fishery surveys

The fishery sector is characterized by a high number of international agencies with specific competence (UE, FAO, OECD, CCAMLR, GFCM, ICCAT, ICES, NAFO, NASCO, IWC, NAMMCO, etc.). They use statistical information for studies or research connected with the preservation and the management of marine resources. One of the most important problems debated in the international working groups is the availability of different and disagreeing statistical sources for the same phenomenon. For this reason Italy has started a process of rationalization and harmonization of the surveys on the fishery and aquaculture sector. The practical outcome of this study was the definition of a sample pilot survey on the catches and the relative values (IREPA & ISTAT 2000) whose objective is satisfying the communitary legislation and, more in general, the national and international informative needs. In two years, it should substitute five surveys currently carried out by ISTAT and the Ministry of the Agricultural and Forest Politics and it should represent the only official statistical source on the sector. A sample survey has been proposed to be able to cope with the problem of the high length of the Italian coast (8.000 km) and the high number of fishery ships (20 thousand).

5. Some innovations in the agricultural sector

Some specific projects have been set up to improve quality of statistics and efficiency of data production processes.

5.1 AGAIN: an integrated system to manage data processing

The issue of improving quality of statistical information has recently been addressed by many national statistical institutions. For this reason Eurostat has issued some guidelines that summarize the leading principles to be followed when carrying out a survey (Eurostat 2000).
Following Eurostat directives and the users' request for information of better quality and within the process of renewing the agricultural statistics sector, the data production process - as already explained - has been completely revised.
One of the emerging activities aiming at improving efficiency of the estimates and the quality of the final information was the setting up of a computerized system (called AGAIN) able to support checks and control procedures and to provide documentation of all the correction, missing data integration and estimation process (Salvi 2000, Benedetti et al. 2001). The advantages of using integrated, user-friendly and flexible systems in the production of statistics are well known. They allow a reduction of the resources devoted to check procedures using macro-editing interactive and automatic techniques. In addition they allow documenting the entire process and providing quality indicators (e.g GEIS developed by Statistics Canada - GEIS development team 1991).
The software architecture has been designed to support the following main purposes: some basic utilities, estimation procedures, imputation and/or integration of missing

total response, data check and control, quality indicators. The last two modules are devoted to get a better quality level by minimal data changes and to report about them. The main features of this tool are portability (since it only requires having SAS license with the module AF) and adaptability. Even though the system has been tested on conjunctural data, coming from diaries and slaughtering houses surveys, it will be adapted in the close future to structural surveys such as the farm structure survey.

5.2 The use of CATI technique

For many years the Italian National Institute of Statistics experienced that a high attrition rate (close to 50%) and a large delay in response affect mail-conducted monthly surveys in the agricultural sector. These issues are due to the fact that statistical units tend to forget their duties of data transmission as much as the time passes from the date of questionnaire reception. As an example, in 1998 the attrition rate in the monthly survey on milk and dairies rose from the 55% in January to the 69% in December with a peak of 84% in the South of Italy at the end of the year. The delay in questionnaire transmission (and in producing the statistical tables for Eurostat) was also quite high reaching sometimes 40/50 days.

To cope with these problems, the sampling technique and the Computer Assisted Telephone Interviewing method has been adopted for surveys on diaries, slaughtering-houses, sowing forecasts and pesticides. The experience was satisfying. In fact, as an example, the attrition rate for the survey on milk and dairy products during year 2000 has always been less than 19%, while the response refusal has been quite negligible (no more than 7 units per month). Also a good monitoring of the archive variations has been made possible.

5.3 Area sampling frame

Conjunctural data are produced, on a monthly basis, on the areas and production of many different agricultural products as required by the European Union. Currently such information does not come from direct observation but is derived from evaluations of a group of experts. These estimations are not comparable with those coming from the survey on farm structure that provides, once a year, sample-based data on the same issue.

With the aim to provide reliable and quickly available statistics on agricultural production, a pilot project has been launched in co-operation with the Ministry for Agriculture and Forestry Politics and with the Consorzio ITA. It will allow to plan a new system of surveys on agriculture based on areal sampling, rationalizing the various data sources existing in this sector. The areal sampling frame uses the land as a mean of detection and selection of statistical units (FAO 1996, FAO 1998). It has many advantages compared to a traditional statistical frame. Mainly, it does not require defining a complete list of the statistical units in the population and allows the use of auxiliary information derived from remote sensing (that also refers to land instead of farms).

The base of the area sampling frame is the partition of the national agricultural area in portions whose borders are designed according to a set of well-defined criteria (total area, homogeneity as for the land use, comparability with administrative areas, easy detection on a map, among others).

Collection of data goes through two steps: the enumerators detect the sampling units on the ground with the help of a detailed map and satellite pictures, then they observe the main agricultural products cultivated in that area. Estimates of productions are derived by integrating surface data with estimation of yields in that portion of land. Information coming from remote sensing is used as auxiliary data to improve the efficiency of final estimates.

5.4 Data capture

Due to the spread of technology in the daily working activity of many public and private institutions, data are frequently available on many different physical supports. The use of e-mail as a mean of quick data transmission and of the web as a direct source of data now gives the opportunity to use sophisticated tools of data capturing. The main advantage in the use of these multifunctional systems consists of a large reduction in the time devoted to the data recording. This is a problem continuously faced by statistical institutions, which need to save time and human resources in their daily activity. For this reason a pilot survey has been planned on a survey so far carried out at the Chambers of Commerce by mail and fax. Currently the Chambers are contacted biannually and fill in a questionnaire containing data on livestocks and animal productions with a large detail of categories. Data, after reception, have to be recorded on a computer and processed to get national estimates. Starting from 2002, to reduce the time of data processing, questionnaires will be captured using specific software such as TELEform. This is a tool that allows to obtain data on different physical platforms, g.e. faxes or computerized questionnaires, with a very low percentage of misreadings. The Chambers will be required to preferably use e-mail to reduce also the time of data transmission. The result of the experiment will obviously depend also on the capability of the statistical units to adapt to the new system and respect the deadline of the survey.

5.5 Data diffusion on the web

In the new agricultural statistical system a specific emphasis it has been placed on customer satisfaction in terms of quality, completeness and timeliness of the information provided. The last requisite, in particular, has been achieved through the increase of the free diffusion form on the WEB that allows to make available information more timely and for the larger number of users. Currently a section in the electronic publication is devoted to the time series of non-structural data on production and external trade of meat and on cultivation. Moreover, some articles on the recent results and comments of the main surveys are available. In the near future a new web site will be set up. It will show the results of the main statistical surveys and will give the opportunity to download data according to user-defined queries.

References

Benedetti (2000): L'impostazione di un sistema integrato di rilevazioni sulle aziende agricole: soluzioni adottate problemi aperti, Atti del seminario *La revisione metodologica delle indagini nel settore primario*, Istat (ed.), Roma, 24 Maggio 2000.

Benedetti R., Martino L & Salvi S. (2001): Un sistema interattivo per la valutazione ed il miglioramento della qualità delle indagini. Paper submitted for the intermediate conference of SIS Society "Processes and statistical methods of evaluation ". Roma, 4-6 2001.

Bethel J. (1989), Sample allocation in multivariate surveys, Survey Methodology, 15, 47-57,

Consorzio Italiano per il Telerilevamento in Agricoltura (2000): AGRIT-2001. Documento interno Consorzio ITA.

Deville J. C., Särndal C. E. (1992), Calibration Estimators in Survey Sampling. Journal of the American Statistical Association , 87, 376-382.

Eurostat (1999): Seminaire FADO. Etude et recherche. Theme 5: Agriculture et peche. Luxemburg: Office des publications officielles des Communautes europeennes.

Eurostat (2000): Budget 2000. Luxemburg: 2000. Published on the web.

Eurostat (2000): Assessment of quality in statistics. Item 4 of the agenda "Standard quality report". Luxemburg 4-5 Aprile 2000

FAO (1996): Multiple frame agricultural surveys. Volume 1: current surveys based on area and list sampling methods. FAO statistical development series n.7. Rome: FAO.

FAO (1998): Multiple frame agricultural surveys. Volume 2: Agricultural surveys programmes based on area frame or dual frame (area and list) sample designs. FAO statistical development series n.10. Rome: FAO.

Filippini R. (1998): Il progetto ASIA-Agricoltura: contenuti e stato di avanzamento. Roma: Documento interno Istat.

GEIS development team (1991): Generalized Edit and Imputation System Operations Users' Guide. Statistics Canada.

IREPA, ISTAT (2000). Proposta di una indagine campionaria per la stima delle qualità e del prezzo medio dei prodotti della pesca sbarcati sul territorio italiano da navi da pesca comunitarie e dell'EFTA, nel corsodi ogni mese di calendario" (Reg. CE n. 1382/91 modificato dal Reg. CE n. 2104/93). Documento interno Istat-Irepa.

ISTAT (2000): Rapporto Annuale, anno 1999. Roma: Istat.

ISTAT (2000): Codice della statistica ufficiale – IV La normativa d'indagine comunitaria, anno 2000. Roma: Istat

ISTAT (2000): Statistica dell'agricoltura. Anno 1997. Capitolo 1: Caratteristiche strutturali e produttive delle aziende agricole, pp.17-82. Roma: Istat.

Martino L. & Greco M. (2000): La Revisione metodologica delle indagini sui prodotti lattiero-caseari, Atti del seminario *La revisione metodologica delle indagini nel settore primario*, Istat (ed.), Roma, 24 Maggio 2000.

Monducci R. (2000): L'utilizzo dei dati amministrativi per la produzione delle informazioni statistiche sui risultati economici delle imprese. Quinta conferenza nazionale di Statistica, Sessione "Integrazione tra registri e rilevazione statistica. Roma: 15-17 Novembre 2000.

Neyman J. (1934): On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97, 558-606.

Salvi S. (2000): Lo sviluppo di sistemi automatici di gestione delle indagini in SAS/Frame, Atti del seminario *La revisione metodologica delle indagini nel settore primario*, Istat (ed.), Roma, 24 Maggio 2000.

Särndal C. E., Swensson B & Wretman J. (1992): Model assisted survey sampling. Springer-Verlag: New York.

Schirinzi G. (1998) Lineamenti per una ristrutturazione del sistema di statistiche agricole in Italia, Documento interno Istat.

A European Approach to Area Frame Surveys

Jacques Delincé[1]
Joint Research Centre of the European Communities (JRC)
Postal Address: TP441, Via Fermi 2, I-21020 Ispra, Italy
Tel.: +39-0332-7855791, e-mail: jacques.delince@cec.eu.int

Abstract: A new area frame survey started this year in the 15 Member States of the European Union. This LUCAS project (Land Use / Cover Area Frame Statistical Survey) adopts a two stage point sampling approach (100,000 points clustered in 10,000 primary sampling units). The survey being multipurpose (agriculture, environment, countryside) a systematic sampling plan (regular grid) has been adopted and efforts have been made to define consistent land use and land cover classification systems. Precision is expected to be around or better than 2% for the main categories. On the basis of a detailed analysis of the results in 2002, the decision will be taken whether or not to repeat the operation in 2003.

Keywords: Agriculture, Area frame, Environment, Land use, Point sampling.

1 The context

The production of agricultural statistics at the European Union level is the result of a close collaboration between Eurostat (the Statistical Office of the European Communities) and the National Statistical Services. Eurostat is in charge of the definition of the characteristics of the surveys (methods, nomenclature, accuracy, timing) and aggregation of the data at EU level (on the base of regulations or gentlemen agreements). Member States are fully responsible for the practical organization related to the data collection and the production of national or regional statistics. More details on this subject can be found in Delincé and Muthmann (1999)
Since 1988, the Commission services (General Direction Agriculture, Eurostat, JRC) have been working on the basis of three Council and European Parliament Decisions, in order to promote new orientations in the structure of EC Agricultural Statistics. The most recent (Decision n°1445/2000/EC of the European Parliament and the Council of 22.05.2000 "on the application of area frame survey and remote sensing techniques to the agricultural statistics for 1999 to 2003") refers mainly to two activities:

- The implementation at Community level of an area frame survey,
- The use remote sensing, in particular with an agro-meteorological system being made operational.

[1] Formerly Eurostat F2 (Statistical Office of the European Communities), postal address: Jean Monnet Building, L-2920 Luxembourg.

Their purpose is to:

- collect the data needed to implement and monitor the Common Agricultural Policy and analyze interactions between agriculture, the environment and the countryside;
- provide estimates of the areas under the principal crops;
- ensure that the condition of crops is monitored until harvesting, so as to enable early estimates of yields and production to be made.

Experience in use of area frames has been gained at the national level (Carfagna et al.,1992, MAPA, 1990, Gay, 2001), at the Commission level (Taylor and al. 1997, Gallego and al. 1994) and at the international level (Cotter and Nealon, 1987). Detailed information on all these projects can be found in the FAO (1998) publication. Building on this experience, Eurostat started in 2000 the implementation of the LUCAS project (Land Use / Cover Area Frame statistical survey) and is carrying it out in 2001 in all 15 Member States.

2 Objectives of the survey

Land cover and land use are of high importance in the definition and evaluation of certain common sectorial policies particularly on the environment, agriculture, and transport, and especially in the integration of those policies based on comprehensive assessment and planning.
The Objectives of the survey have been defined as follows:

- To obtain **harmonized data** (unbiased estimates) at EU 15 level of the main Land Use / Cover areas and changes. Precision is expected to be around or better than 2% for main categories such as wheat, cereals, arable land, permanent grassland, permanent crops, forests, urban areas, inland waters.
- To extend the **scope of the survey**, beyond the usual agricultural domain, to include also aspects linked to environment, multi-functionality, landscape and sustainable development.
- To offer a **common sampling base** (frame, nomenclature, data treatment) that interested member states can use to obtain representative data at national/regional level by increasing the sampling rate, respecting the general LUCAS approach.
- To evaluate the strengths and weaknesses of a **point area frame survey** as one of the pillars of the future Agriculture Statistical System (area frame means that the observation units are territorial subdivisions instead of agricultural holdings as in the Farm Structure Survey).

LUCAS consists of 2 phases:

- during phase 1, data on land cover / land use and environmental features are collected in the field in the spring time at around 100,000 observation points in Europe;
- phase 2 is concerned with interviewing about 5,000 farmers in autumn to obtain additional technical or environmental information.

3 Why a systematic area frame point sampling?

The choice of an area frame instead of a list frame is easy to understand in the case of a land use / land cover survey having the whole of the EU territory as the target population target. Lists are difficult to elaborate and maintain up to date. Not even the national cadastral databases would have provided a reliable list of land managers.
More open were the questions related to the details of the sampling plan (stratification or clustering) and to the size and nature of the units sampled (points, lines or areas).

3.1 Sampling plan

The reduction of the variance of the estimator in finite population modeling can be obtained in several ways among which stratification and clustering are the most in use.
In the case of a multipurpose survey, detailed stratification is questionable as it necessitates prioritizing the nomenclature classes and favoring some to the detriment of others. Even if the CORINE land cover map had been fully available at EU level (EEA, 1999), its use would not have had a noticeable influence on variances without an increase/decrease of sampling fractions in function of the interest of the different strata. Such a choice would have been obtained to the detriment of the multipurpose objective and would have limited the flexibility of the survey to answer the future additional demands deriving from development of EU policies. For this reason, the only stratification kept was the geographic delimitation of the 15 EU Member States.
On the contrary, the adoption of a clustering approach through a systematic plan allows the goal of reduction of variance to be met, while keeping the flexibility and the multipurpose aspects. The amount of variance reduction depends on the spatial correlation of the underlying stochastic processes and can be evaluated. Having access to the data from the 1998 TERUTI survey (SCEES, 2001), we calculated the variances of all the land use categories of this survey covering the whole of France (550,000 Km²) in two ways: considering it firstly as a random and secondly as a systematic two stage survey. We compared both approaches by calculating a regression line between the variances and the totals (1.000 ha) of the 88 land use/cover classes.

Table 1: *Linear regression between variance and area coverage of the 88 land classes of the TerUti 1998 survey, in case of random and of systematic sampling.*

Model	Intercept	Slope	R^2
random	-30.8	0.674	0.63
systematic	-59.7	0.398	0.88

This shows an average relative efficiency Er = 1.7 of systematic sampling, in the same order of magnitude than the average stratification efficiency of 1.5 obtained in the MARS Regional Inventories (Taylor and al., 1995)

3.2 Segment size

The principle of keeping a segment of a certain size was never questioned. For reasons of economy, it is obvious that when a surveyor drives half an hour to reach a certain place, more than one land entity should be surveyed in order to optimize the costs. Using the TERUTI data where the segments are 6 km apart and the segments covering 324 ha (in fact 6x6 points 300 m. apart), we computed the slope of the regression line liking the variance of the area of the different land use classes for different sampling strategies.

Table 2: *Slope of the linear regression relating the Estimation variance and the extent of the class of interest for different sampling strategies*

Grid distance	Pattern of points inside segment				
	6x6 324 ha	3x3 external 324 ha	3x3 centered 81 ha	2x6 108 ha	2x5 90 ha
6x6 Km	0.40	0.50	0.74	0.62	0.68
12x6 Km	0.96				
12x12 Km	2.41			3.51	3.82
18x18 Km	5.36			7.23	7.88

From this table we conclude that:

- when the distance between segments increases from 6 to 12 Km, the variance of the estimates increases more than proportionally, showing that a positive correlation exist at that scale (in absence of spatial correlation one would have expected that the 12x12 Km coefficient would have been 4 times greater than the 6x6 Km one (so 1.6 instead of 2.41 in case of 36 points per segment). For distances above 12 Km, the absence of correlation seems to have been reached (5.36 < 2.41x2.25=5.42),
- when keeping the size of the segment to 324 ha, the reduction of the number of points by a factor 4 increases the variance by only 25% (due to the fact that more than 90% of the variability is between segments). This suggests reducing the number of points.
- when keeping only 9 points but on a smaller segment size (81 ha instead of 324) , the variance increase significantly due to the local positive spatial correlation (0.5 -> 0.74 for the same sample size). Anyway, segments of 81 ha are more efficient than the 324

one from the cost (3 Km to walk instead of 12) and the variance point of view (0.74<4x0.4).

- when reducing the sample size from 6x6 to 2x5 points, the variance increases by a factor of 1.7 (compared to 3.6 for the sample size). The shape of the segment does not appear to be of major importance, as the gain of efficiency between the square segment of 81 ha compared to the rectangular of 90 ha is only 2% (1.111/1.088). The shapes being equivalent, we prefer a pattern of 2 series of 5 points, due to the fact that, with the 3x3 pattern, the surveyor has one kilometer more to walk (1.27 km instead of 0.3) when coming back from the last to the first point.

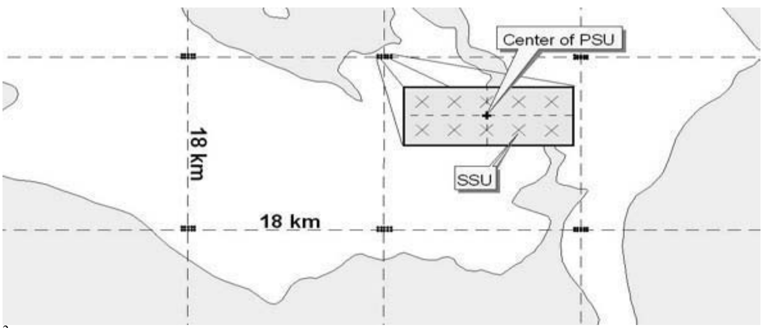
In addition to these calculations, we used the Neyman optimal allocation formula (Cochran, 1977) to compute the optimum number of points per segment taking into account the structure of costs and variance. Considering that the fixed costs were around 49 Euro and that the variable costs were 14 Euro at segment level and 2.7 Euro at point level, the optimized number of points per segment varied from 3 for forest areas to 10 for urban areas.

3.3 Area or point sampling

The decision to sub sample the segment area through a grid of points is straightforward when looking to the structure of variance (variance between segments and variance between points within segments). In all our sampling simulation, the component of variance related to the second stage rarely reached 5% of the total variance.

For the above reasons, the sampling plan adopted consists of one systematic square grid[2] (18x18 Km) and rectangular segments of 2x5 points 300m apart (90ha). This imposes -due to the absence of replicates- an approximate estimation of variance through the differences between neighboring sampling units (Cochran, 1977).

Figure 1: *LUCAS sampling design*



4 Other technical aspects

[2] In order to be compatible with existing national frames, two exceptions have to be mentioned for Spain and Italy where a distance of 250 m has been retained between points. In addition, the grid used in Spain is a 20x20 Km.

4.1 Organizational aspects

The LUCAS project is financed by the EU General Direction Agriculture for an amount of 3 MEUR in 2001. **Eurostat** is responsible for all methodological issues and project management, with the technical support of the **Space Applications Institute** (DG JRC).
On the base of a call for tender published in June 2000 in the EU Official Journal, contracts were attributed in January 2001 to realize the work. The documentation and the quality assurance of the project have been entrusted to the **CESD Communautaire** and carrying out the LUCAS surveys to the following contractors in the 15 EU Member States:

Belgium & Luxembourg	G. I. M.
Denmark	KAMPSAX
Germany	EFTAS
Greece	GEOAPIKONISIS
France	**SCEES**
Spain	**AGROSEGUROS**
Ireland & United Kingdom	**ADAS Consulting LTD**
Italy	**Consorzio ITA**
The Netherlands	**SYNOPTICS**
Austria	GEOSPACE
Portugal	TERRACARTA
Finland	**Inf. Center, Min. of Agr.and For.**
Sweden	**STATISTICS SWEDEN**

It should be noted that 8 of the 13 country contractors are (or are close to) the national structures in charge of Agricultural Statistics.

4.2 Expected precision and national enhancements

On the basis of the available data, Eurostat optimized the sampling plan and assessed in advance the expected precision of the LUCAS survey. For the sampling plan adopted, the relation between the land cover area (X, 1000 ha) and the variance of estimation (Y) is linear:

$$Y = -1{,}056 + 7.89\,X \quad (R^2=0.95)$$

The coefficients of variation at European level for the main Land Use/Cover classes are presented in table 3.
The success of the LUCAS project will be based on the results obtained (precision, reliability, timeliness) but also on the willingness of the Member States to adopt the area frame methodology. France, Spain and Italy have already a long tradition of using area frames to produce agricultural statistics and LUCAS will benefit from an increased sample size due to national investments. It is expected that in 2001 results will be available with a sample size increased by a factor 9 in France and a factor 12 in Spain. In Italy, a factor 10

should be reached in 2003. This will, for example, lower the relative precision on the area of cereals below the value of 1%.

Table 3: *Expected Coefficient of variation at European level*

Land Cover	CV	Area (1,000ha)
U.A.A.	0.75	140,619
Arable	1.02	76,368
Permanent Grass	1.22	52,611
Permanent Crops	2.60	11,574
Cereals	1.45	37,343
Wheat	2.13	17,204
Fallow & Gr.Manure	2.80	9,932
Durum wheat	4.98	3,046
Grain Maize	4.45	3,852
Dried pulses	6.23	1,888
Potatoes	7.13	1,403
Sugar beet	5.94	2,089
Rape seeds	5.36	2,601
Sunflower	5.09	2,907
Soya beans	11.74	358
Cotton seeds	11.28	424
Inland waters	2.63	11,279

4.3 Documentation

A particular effort has been placed on the documentation of the LUCAS project. Technical documents have been created to cover the following topics:

LAND/LUCAS/1	The sampling design
LAND/LUCAS/2	The nomenclature
LAND/LUCAS/3	The geometric requirements
LAND/LUCAS/4	The instructions for surveyors
LAND/LUCAS/5	The farmers interviews (LUCAS Phase 2)
LAND/LUCAS/6	The database
LAND/LUCAS/7	Quality assurance and control procedures
LAND/LUCAS/8	The structure of reports
LAND/LUCAS/9	The estimation method

They can be found and downloaded from the CIRCA web site:
http://forum.europa.eu.int/Public/irc/dsis/Home/main

4.4 Nomenclature

The design of the LUCAS nomenclature is based on the results of an Eurostat project on a “Classification system for land cover and land use” (Eurostat 2000). Various existing national and international classification systems have been analyzed to establish best practice for the construction of a classification system to be used in the framework of LUCAS. In close collaboration with an international expert group of official statisticians from Portugal, France, Germany, Switzerland and Slovenia a classification system has been proposed. Agreement has been reached and the classification system adopted at the Eurostat Working Group on Land Use Statistics.

The main points which were the object of analysis and choices were:

- the definition of Land (extended to inland water), of Land Cover (above the earth's surface) and of Land use (socio-economic activities), special attention has been given to a clear separation between the concepts of "use" and "cover",
- the observation unit: a circle of 3 m of diameter except in case of heterogeneous areas or permanent crops,
- the separate registration of multiple use/cover in order to avoid mixed classes,
- the treatment of points on limits and the use of ortho photography,
- the documentation of the classes in terms of definition, list of inclusions and exclusions, compatibility between land cover and land use,
- the correspondence with existing national or international classification systems.

This resulted in a Land Cover classification of 57 classes regrouped in 3 hierarchical levels and in a Land Use nomenclature of 14 classes.

4.5 Environmental features

LUCAS was set up primarily to collect information about land cover/land use, with a specific focus on agriculture. However, this survey provides a unique opportunity to collect additional information within an environmental context as well.
The uniqueness refers to the:

- Systematic spatial coverage, statistically representative for the entire EU territory,
- Regular intervals of field observations at a yearly basis,
- Direct observation at field level,
- Potential link of land cover/land use to environmental issues.

It has thus been decided to collect environmental information on the first five points of the LUCAS segment considering that they form a linear transect of 1.2 Km.
On this sub sample, information will be collected on the presence of irrigation, isolated trees, natural hazards, erosion and noise. In addition, digital photos will be collected in order to create a European database of landscape and countryside information.
The landscape diversity will also be described by estimation of the length of linear features (by counting the number of intersections per category) and the average dimension of the main land cover patches.

4.6 Estimators

Three main types of estimators will be used in function of the variable of interest:

- Area estimates: point sampling has the characteristic of having a sampling probability proportional to the size of the selected unit. For area estimates of land cover and land use classes, this has the major interest that an unbiased estimate is obtained by multiplication of the area of interest by the percentage of points falling in the category of interest.
- Length estimates: line intersect methods (De Vries, 1979) allows an unbiased estimation of relative density of linear elements (without constraints on their shape) as long as they are randomly and independently distributed. The only information needed in that case is the number of intersections.
- Phase 2 estimates: for the variables contained in the farmer interviews, the extrapolation factor is proportional to the sampling probability. This implies collecting the total area of arable land managed by the selected farmers.

5 Conclusions

The LUCAS survey has already been a long process, started in early 1998 and finally adopted at European Parliament and Council level in June 2000. An important time has been allocated in the years 2000 and 2001 to the survey preparation in term of work definition, selection of contractors, documentation and follow-up.
It is hoped that detailed results will be obtained in June and November 2001. These will allow the potential of the area frame approach and of the options taken to be evaluated. Due to the huge amount of information available, most of the data analysis will take place in the year 2002 and will be dedicated:

- to the complementary analysis of the results (alternative modeling of sampling, photo interpretation as alternative to field visit, database elaboration),
- to the extension of the survey to three candidate countries (Estonia, Poland, Slovenia),
- and to internal EU discussions on the appropriateness of a renewal of the operation in the year 2003. This will be of particular importance to show the potential of the area frame approach for the estimation of land use changes.

We would like to thanks here the numerous people that have been and are still involved in the LUCAS Project; in particular the members of the Eurostat Working Group on Land Use Statistics and the LUCAS contractors. A special mention has to be done of the French Ministry of Agriculture (SCEES) for providing the data on which the LUCAS sampling optimization has been based.

References

Carfagna E., Ragni P., Rossi L., Terpessi C. (1992). *Area Frame: Un nuovo strumento per la realizzazione delle stasistiche agricole italiane*. Atti statistica per le analisi territoriali e spaziali. Univ. Parma, pp. 25-36.

Cochran W.G. (1977) Sampling Techniques. Wiley 428 pp.

Cotter J., Nealon J. (1987). *Area Frame design for Agricultural Surveys*. U.S. Dept. of Agriculture. Nat. Agr. Stat. Serv.

Delincé J. & Muthmann R. (1999) *Parallel Responsibilities of International and National Statistical Offices and Ministries of Agriculture, their Co-operation and Solving Common Problems.* Meeting on Food and Agricultural Statistics in Europa, Conference of European Statisticians, Geneva, June 1999, paper n° GE-99-31347.

De Vries P.G. (1979) Line intersect sampling. In: *Sampling Biological Populations* (Cormack, Patil, Robson ed.), Int. Coop. Publish. House, Burtonsville, 1-70.

EEA & ETC/Land Cover (1999). *CORINE Land Cover Technical Guide*. At http://etc.satellitus.se/the data/Technical Guide/idex.htm.

EUROSTAT (2001). *Manual of Concepts on Land Cover and Land Use Information Systems*. Office for Publications of the EC, Luxembourg, theme 5: Agriculture and fisheries, Collection: Methods and nomenclatures, 106pp.

FAO, 1998, *Multiple frame agricultural surveys, Volume2: Agricultural survey programmes based on area frame or dual frame sample designs*. FAO statistical development series, n. 10, 242 pp.

Gallego F.J., Delincé J., Carfagna E.(1994) *Two-Stage Area Frame Sampling on Square Segments for Farm Surveys*. Survey Methodology , vol. 20, No. 2 , pp. 107-115.

Gay C., Porchier, J.P. (1998) *Land Cover and Land Use Classification using TER-UTI* , Agricultural Statistics 2000, Washington, March, 18-20. Pp193-201.

MAPA (1990) *El Marco de Areas como Instrumento de base para la Estadística de superficies de cultivo*. Boletín Mensual de Estad. Min. Agric. n. 12, pp.85-106. Madrid

SCEES (2001) *AGRESTE: Chiffres et Données Agriculture*. N°133, mars 2001, auteur: V. Rabaud, Paris, France.

Taylor J., Sannier C., Delincé J, Gallego F.J. (1997). *Regional Crop Inventories in Europe Assisted by Remote Sensing: 1988-1993.* Synthesis Report . EUR 17319 EN, Office for Publications of the EC. Luxembourg. 71 pp.

6 June 2001

Invited Paper Session

DATA INTEGRATION ISSUES

Chair: U. Jorner

Multiple Use of Data Within the System of Agricultural Statistics in the Federal Republic of Germany

Martin Schmidt
Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft
Postfach 14 02 70, 53107 Bonn, Germany
e-mail: martin.schmidt@bmvel.bund.de

Hannelore Pöschl
Statistisches Bundesamt, Zweigstelle Bonn
Postfach 17 03 77, 53029 Bonn, Germany
e-mail: hannelore.poeschl@statistik-bund.de

Keywords: Integrated survey, unification of cut-off limits and sampling plans, administrative data, Special Harvest Assessment

1. Introduction

In the Federal Republic of Germany, statistics on holdings and production in agriculture and forestry as well as in the fisheries sector form a well coordinated system, which allows a continuing monitoring of farms and production. In this process, the collection and processing of agricultural statistical information is on the one hand under pressure to meet new demands expressed by agricultural policy makers at national and international levels, and to rapidly make the results generally available without charge, if possible. On the other hand the costs for data collection are to be reduced. As a result of these partly conflicting demands it is necessary to use data already available as comprehensively as possible for the purposes of agricultural statistics and to adjust schemes, methods and processes in a way that the integration of information can be realised smoothly and the results achieved this way reflect the facts to be described correctly.

The official German statistics – at least in the case of some special agricultural statistic surveys – has already been engaged in this process for a longer time. In 1998, however, a decisive milestone was reached with the amendment of the national Agricultural Statistics Act. This Act was preceded by comprehensive methodological studies to review and modify concepts of collection, processing, tabulation and presentation. Since then the further development of the new system to be described in the following is the focus of methodological activities. At the same time work towards validating the previous decisions is continuing.

2. Integration of production statistics and structure statistics

2.1 Integrated survey

The organisational concept of simultaneously collecting characteristics on farm structure and production practised since 1999 is called “integrated survey“ in German

agricultural statistics. To this end, the production statistics "animal census" and "land use survey" are integrated in the farm structure survey; the same applies to the special surveys covering wine and horticulture.
This proceeding is based on a homogeneous group of questionees. Hence the units defined in the Agricultural Statistics Act were farm and forestry holdings. Holdings in the sense of this Act are economic units subject to a uniform farm management and producing agricultural or forestry products. In Germany, this restriction of respondents is possible, since the production potential of areas and livestock is almost exclusively concentrated on agricultural and forestry holdings. In addition, higher cut-off limits were introduced in order to simultaneously reduce the burden on information providers (see table 1).
Based on analyses of the previous surveys on farm structure, land use and livestock, the consequence of the harmonisation and increase of lower cut-off limits was that out of the approximately 700,000 collection units (status: 1995) in the field of agriculture

- the work load required to establish agricultural statistical surveys of some 520,000 agricultural holdings still to be questioned has been reduced;

- some 180,000 small agricultural holdings (< 2 ha UAA), cultivators of individual areas and animal keepers who do not qualify as managing a holding were fully exempt from the obligation to provide information.

In the year 2000 a total of just under 460,000 agricultural holdings were included in the census in accordance with the new delimitation.
The loss of information on the production potential was considerably smaller so that the change is an acceptable compromise for the main users. In the „integrated survey" the utilised agricultural areas decreased by less than 1 per cent compared to the previous surveys (1995 status), the number of bovine animals and pigs fell by just over 1 per cent, with variations according to the individual crop species and livestock categories:

Utilised agricultural area	- 0.8 %	Bovine animals	- 1.2 %
spec: arable land	- 0.3 %	spec: dairy cows	- 0.8 %
permanent grassland	- 1.7 %	pigs	- 1.2 %
land under vines	- 2.6 %	spec: breeding sows	- 0.8 %
fruit plantations	- 1.9 %	fattening pigs	- 1.5 %
		chickens	- 4.6 %

Since the structural change continued between the 1995 analysis and the change in the year 1999 – cessation of production of small holdings, increase of area in larger holdings – the actual difference in land and livestock has probably been even lower.

2.2 Unification of the sampling plans
A uniform sampling plan covering up to 100,000 holdings is also intended to reduce the response burden. When drafting the plan it was tried to meet the requirements both of structure statistics and production statistics. Hence a stratified procedure was chosen. In terms of regional aspects, the stratification was geared to the *Laender*. In terms of technical aspects, the stratification characteristics and stratification limits were exclusively determined according to criteria for sampling methods. The stratification

according to size classes of the utilised agricultural area (UAA) serves as an approach for a pre-selection grouping of holdings, which is traditionally stable and effective. Moreover, additional strata were introduced in order to increase the accuracy of results. They include the relatively small number of holdings standing out against the majority of agricultural holdings because of the one-sided orientation or the special significance of their production. This approach guarantees that these holdings are covered with adequate precision, and on the other hand it is ensured that within the size classes of the UAA the holdings form largely homogeneous groups with regard to the characteristics covered. The strata limits are determined with the help of the pre-period results of the census. Also, selection intervals for the strata delimited in accordance with technical and methodological criteria are determined for each *Land* on the basis of the mean value and variance of the information on the census provided by the individual holdings. The current farm register serves as a basis for the selection (see section 3.2). It stores all relevant data characterising agricultural and forestry holdings. The allocation of the holdings to the strata is based on allocating each holding to the first stratum which meets its criteria. This is done independent of whether it also fulfils the criteria of other strata.

2.3 Survey programme and multiple use of data

The survey programmes of the farm structure survey and the production surveys are coordinated in a way that for regularly recurrent enquiries all required farm data (see table 2) are only requested once and then inserted in the corresponding tabulation. Thus the information providers are only questioned once about the respective complexes of characteristics, yet the information provided is included in multiple ways in different tabulation processes.

In accordance with the scope of the characteristics the results are used in various ways:

- Areas broken down into principal uses and crops, plant groups and plant species;
- Livestock broken down into animal species and categories, used also as basic parameter of meat production forecasts;
- Employment and working hours;
- Number of holdings, broken down according to size, legal form, production focus, type of farming;
- Socio-economic conditions, ownership and tenant relations, vocational training, transfer of farm (successorship), renting of rooms and other facts.

In addition, the data will be included in the

- Production statistics of field crops (cultivated areas),
- Supply balance sheets (usable production),
- Agricultural accounts (production value, intermediate consumption, labour input) and
- Environmental indicators.

The system also provides the advantage that all characteristics of a farm can be linked with other information on the farms.

2.4 Further expansions

In a further step to compose agricultural statistics, the horticultural survey scheduled for 2005 is to be included in the integrated survey. This means that data on the farm structure and employment as well as on land use will be taken from the simultaneously conducted farm structure survey (including land use survey). In addition, characteristics of the survey on ornamental plants and nurseries are to be integrated so that there is only one additional enquiry instead of three individual surveys. This procedure is now being conclusively discussed by experts and will be reflected in the amendment of the Agricultural Statistics Act scheduled for 2002.

3. Use of other sources for statistical purposes

Considerable savings were realised and some burden was taken from the information providers thanks to the introduction of the integrated survey in 1999 by
- combining surveys
- increasing and harmonising lower cut-off limits
- extending survey periodicities
- reducing the survey programme/streamlining catalogues of characteristics
- having a sample survey instead of a census for individual characteristics.

Additional effects are achieved by using administrative data and external address registers. In this process it has to be considered that administrative data are collected for specific facts and that the delimitation of collected data and the used definitions of stored characteristics do not always coincide with those of statistics. Nevertheless administrative data have successfully been used for official agricultural statistics. This is possible because the statistical offices were able to solve at least part of the technical problems through methodological work. A corresponding legal provision of the Agricultural Statistics Act determines the type and scope of use.

It must also be noted that the German privacy protection law unilaterally allows only the use of administrative data for statistical purposes. By contrast, individual data from statistical surveys may exclusively be used for statistical purposes, i.e. that in the course of implementing subsidies they may not be used simultaneously by administrative authorities.

3.1 Administrative data

In the Federal Republic of Germany a data source used for agricultural statistics in administrative enforcement is the EU's , "Integrated Administration and Control System" (IACS). These are administrative data collected in connection with the agricultural policy reform adopted in 1992 by the European Council of Ministers. In order to manage and monitor area-based and livestock-based subsidies, farm data are collected which are also subject to surveys under the Agricultural Statistics Act. In Germany the *Laender* are responsible for implementing and enforcing IACS, and as a consequence there is no uniform application form throughout Germany. The prerequisites for using IACS data for the purposes of agricultural statistics vary considerably in the individual *Laender*. In addition, the administration has only data on those farms and can only make available data on those farms which have submitted an aid request. Information about the remaining farms must still be retrieved directly from the farmers. In addition, the data not included in IACS must be collected through other reporting channels, i.e. by directly asking the persons obliged to provide information. Despite these difficulties, this procedure has meanwhile been used in 6 out of 13 territorial states. In this process, IACS data are inserted in the land use survey (information on the cultivation of arable land) as well as in the livestock survey.

In 1999 it was also possible to use administrative data to a considerable extent for the viticulture survey realised every 10 years. Data on the structure of wine-growing holdings were taken from the agricultural census carried out simultaneously, data on the structure of cultivation and the vine variety were taken from the Community Vineyard Register established in the framework of the Common Market Organisation in wine. The areas under vines available in the vineyard register had to be processed on a per-holding basis. A study examining the feasibility of this approach revealed that the EU Vineyard Register and the official statistics cover a significantly different number of units. This can mainly be attributed to the fact that the agricultural statistics only document technical-economic units (holdings) whilst the EU Vineyard Register also documents property units where the property holder is not identical with the vineyard manager. In order to achieve comparable results, the units from the vineyard register had to be pooled into holdings in the sense of statistics.

3.2 External address registers

In Germany, a uniform farm register is kept for the preparation, management and processing of agricultural statistics. The "farm register" is administered and updated decentrally by the Statistical Offices of the *Laender*. The information on farms stored there is used for methodological and organisational work, inter alia:

- Identification and establishment of sampling units,
- Drawing of random samples for sample surveys,
- Setting up of rotation plans,
- Limitation of the burden on questionees,
- Mailing of survey documents,
- Reception check,
- Consitency-checks from the questionees,
- Realisation of surveys by updating procedures
- Verifying the accuracy of the results,
- Adjustments of samples.

For the updating of the register data the agricultural employer's liability insurance association, a part of social security for farmers, transmits data on the farms every two years. Thanks to this comparison it is possible to detect additional farms which fulfil the obligation to provide statistical information, but which have hitherto not been covered by the Statistical Offices of the *Laender*.

3.3 Prospects

Currently data from the so-called "animal tracing and information system" (Herkunfts- und Informationssystem für Tiere, HIT), which up to now only contains data on bovine animals, are being used in a test run for the survey on herds. This system of labelling and registering bovine animals was set up as a consequence of the first BSE crisis in 1996 and committed the member states to establishing a corresponding database before the year 2000. The central database was established in the Bavarian Ministry of Agriculture by order of the *Laender*. It primarily serves the control of animal diseases, but also the enforcement of EU subsidies for bovine animals. A first test of this database for agricultural statistical purposes was the insertion of HIT data for the livestock survey in two *Laender*. It is intended to take greater advantage of this possibility in the future. In addition to further methodological studies and legal provisions this also requires the approval of the European Commission. Moreover, an examination will have to be carried out as to how far the new pig database to be established at EU level can be used for official statistics.

4. Multiple use of statistical information: examples from special surveys

In addition to the combination of agricultural basic statistics, strong relief and utilisation effects can be achieved by the specific collection of production and marketing data in individual product areas.

4.1 Special Harvest Assessment

The purpose of the Special Harvest Assessment is to provide exact data on the quantity and quality of the harvest of selected crops for the entire German territory and the *Laender* in connection with the land use survey at the earliest possible date. In this process, the final hectare yields of cereals (and potatoes) are ascertained in a sample procedure on randomly selected fields. In addition to the established weight of the harvest quantity which is the basis for the extrapolated hectare yields, the following factors are ascertained in the agricultural testing stations:

- moisture content and
- impurities (black dockage, sprouted grains) in the harvested material.

One partial sample of each of the bread grain varieties of wheat and rye is sent to the Federal Centre for Cereals Research for further tests of general characteristics (see diagram):

a) Quality

The observations concerning quality including the baking qualities cover the following parameters, among others

- groups of Besatz
- protein content
- sedimentation index
- falling number.

b) Contamination

The analyses include in particular

- heavy metals
- residues of plant protection products and
- mycotoxins.

In view of the discussion on the quality and safety of food these results are becoming increasingly important. Meanwhile the statutory requirements have been set to facilitate the implementation of the evaluations for other grain species too, if required.

The measuring method for determining cereal yields and the ensuing tests require a relatively large amount of financing, yet this is justified by the high degree of data accuracy and the diversity of uses. In the sensitive areas of quality and contamination,

the tests at a central institution and in accordance with a uniform methodology allow objective and representative statements.
The areas of use can be summarised as follows:

1. Hectare yields (multiplied by the cultivated area)	- Cereals production - Supply balance sheets for cereals (resources) - Agricultural accounts (production value)
2. Cereal varieties	- Distribution in accordance with the cultivated area
3. Quality parameters	- Quality of grain harvest (situation and chronological comparison)
4. Pollutants	- Pollution situation (also in chronological comparison) - Environmental aspects - Decision aids for the determination of permissible maximum residue levels

According to our knowledge this procedure is unique in the European Union and maybe even world-wide.

4.2 Slaughter statistics
In accordance with the provisions of the Meat Hygiene Act, the official veterinarians and meat inspectors must record the number of animals examined. In this process a distinction is made
- between animal species (bovine animals, pigs, sheep, goats and horses)
- in the case of bovine animals also between use categories (calves, heifers, cows, oxen, bulls)
- between commercial slaughterings and domestic slaughterings and
- according to the origin of slaughter animals (Germany, other countries).

The purpose of the tests is to establish the edibility (fitness) of the meat and its approval for human consumption. The meat hygiene statistics are established on the basis of these results. In addition, the data on the number of slaughtered animals also form the basis of the slaughter statistics.
On the basis of the Livestock and Meat Act and its implementing ordinances the wholesale slaughterhouses and meat products factories must notify prices. Since the number and weight of the animals slaughtered there is also notified, the average slaughter weight of animals of German origin can be derived from these data.
The multiplication of the number of slaughtered animals and the average slaughter weight is the meat production, classified according to types of meat. These data are also included in the supply balance sheets and in the agricultural accounts.
Thanks to the statutorily required meat inspection, coverage is very high. Other arguments for the comprehensive use of these data in statistics are that

- more than 96 per cent of bovine animal slaughterings and almost 98 per cent of pig slaughterings are commercial slaughterings (only in the case of the less significant sheep slaughterings the percentage of home slaughterings stands at a relatively high 65 per cent) and
- in the case of commercial slaughterings some 73 per cent of bovine animals and 78 per cent of pigs are marketed via wholesale slaughterhouses and meat product factories.

4.3 Food industry statistics

In the sectors of cereals, feedingstuffs, sugar, oils and fats as well as milk and milk products, commercial and industrial businesses must submit notifications on the basis of a regulation. Large and medium-sized enterprises notify once a month, smaller businesses with less impact on the market provide their information once every six months.

These products
- have strongly regulated markets in the EU,
- can generally only be consumed in a processed condition,
- are marketed through only a few distribution channels in which
- relatively few companies („bottlenecks“) operate.

The characteristics collected by the food industry are:
- the raw materials supplied by agriculture,
- the quantities of raw materials used in processing,
- the products manufactured and sold,
- the stocks of products and
- in the case of milk also the prices paid to the producers.

The benefit of this survey system is that for some characteristics instead of asking many farmers only a relatively small number of their trading partners need to be questioned (some 3,000 in these sectors taking a cut-off limit into account). In addition, there is a broad range of possibilities to use the data. The results
- are used for the short-term observation of economic trends in these economic sectors and for market reporting;
- represent 100 per cent of the sugar and oils production and an essential part of the milk production (94 per cent of the milk is supplied to dairies);
- are integrated into the agricultural accounts (agricultural sales, intermediate consumption) and
- in the supply balance sheets (sales, changes in stocks, domestic uses);
- form an important database at low regional level for the purposes of food security in the event of emergencies, for example natural disasters.

5. Outlook

The concept of linking production and structure statistics so closely that there is only a homogeneous group of respondents left means a strong boost to efficiency in the official agricultural statistics. This can, however, only be realised if the production potential in terms of areas and livestock is mainly located in medium-sized and larger holdings, such as in Germany, that is that the agricultural production in private households and microbusinesses is relatively insignificant. To this effect such a system cannot simply be transferred to other countries, but is dependent on the production structure of each country.

At least in the EU, the intensified use of administrative data also provides rationalisation opportunities, requiring a further adjustment of methods and concepts. It will therefore be necessary not only to adjust national statistics laws as currently done in Germany, but also to pay more attention to the congruence of statistics requirements and administrative data in the updating or first-time formulation of EU rules. The better definitions and delimitation criteria are coordinated with each other, the more successful it will be to take some burden from the information providers, collect data cost-effectively and provide all required information timely.

Table 1: *Determination of lower cut-off limits of agricultural holdings*

Characteristic	Unit	Distinction made until 1998			No distinction as of 1999
		Farm structure survey	Land use survey i.a.[1)]	Livestock survey	Farm structure survey new [2)]
Utilised Agricultural Area (UAA)	Hectare	1	1/< 1[1)]	1	2
or livestock					
Bovine animals	Heads	8	-	1	8
Pigs	Heads	8	-	-	8
- Breeding pigs	Heads	-	-	1	-
- Other pigs	Heads	-	-	3	-
Sheep	Heads	50	-	3	20
Horses	Heads	-	-	2	-
Poultry (per species)	Heads	-	-	20	-
- Laying hens	Heads	200	-	-	200
- Young hens	Heads	200	-	-	200
- Roosters for slaughter and fryers, broilers and other roosters	Heads	200	-	-	200
- Geese, ducks, turkeys	Heads	200	-	-	200
or special crops					
Area under vines, fruit plantations, hop, tobacco, nurseries, field vegetables	Are	30	< 30[1)]	-	30
Outdoor cultivation of flowers and ornamental plants	Are	10	< 10[1)]	-	30
Medicinal plants, culinary plants, horticultural seeds	Are	1	< 1[1)]	-	30
Vegetables, flowers and ornamental plants grown under glass	Are	1	< 1[1)]	-	3

Note: If the holding fulfils one of the conditions, all survey characteristics must be indicated independent of whether limits are reached.

1) Additional coverage of areas (in particular sales cultivation) below 1 hectare, 30, 10 or 1 are in the land use survey just as special surveys in horticulture, fruit cultivation and viticulture.

2) Including the land use survey, livestock survey and special surveys.

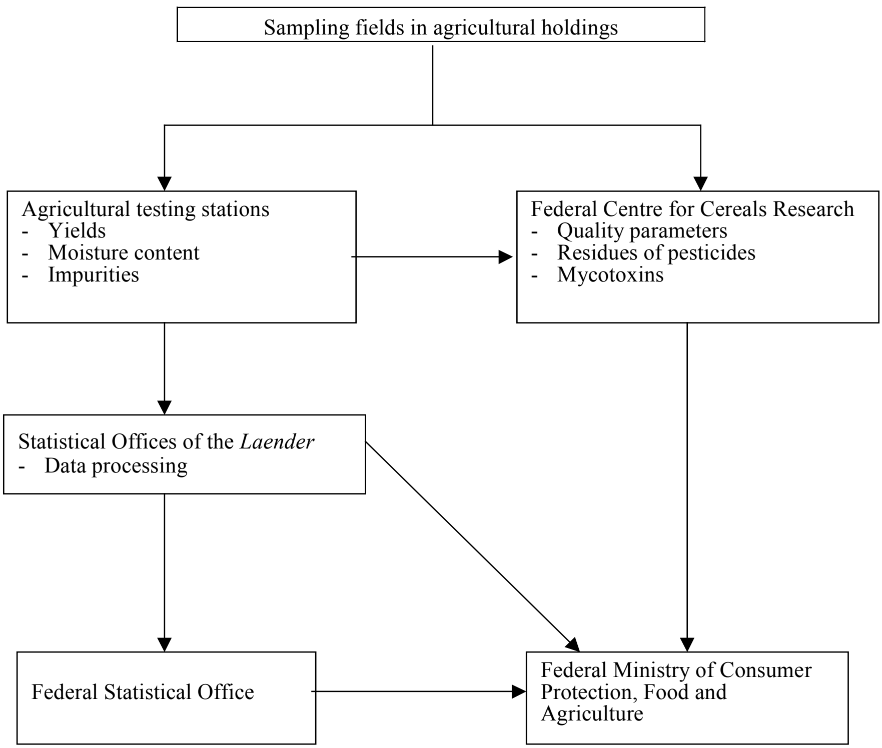
Table 2: *Type and periodic intervals of regular enquiries in the agricultural statistics of the Federal Republic of Germany*

Title of surve	**Date**	**Current example**	**Type of survey**
Farm structure survey (agricultural census) + Land use survey + Livestock survey	May t_1	1999	Census + Census + Census
Livestock survey	November t_1	1999	Sample survey
Land use survey + Livestock survey	May t_2	2000	Sample survey + Sample survey
Livestock survey	November t_2	2000	Sample survey
Farm structure survey + Land use survey + Livestock survey	May t_3	2001	Sample survey + Census + Census
Livestock survey	November t_3	2001	Sample survey
Land use survey + Livestock survey	May t_4	2002	Sample survey + Sample survey
Livestock survey	November t_4	2002	Sample survey

Diagram

Multiple Use of Statistical Data

Special Harvest Assessment for cereals in Germany



Results			
Hectare yields Harvest quantity	Distribution of cereal varieties	Harvest qualitiy	Contamination
Moisture content			

The Implementation of an Information System on Administrative Statistics in Agriculture – Example of One Member State of the European Union

Jean-Claude Porchier
SCEES, Ministère de l'Agriculture et de la Pêche, 251 rue de Vaugirard
75732 PARIS Cedex 15, France
e-mail: jean-claude.porchier@agriculture.gouv.fr

Abstract: Administrative data systems and statistics systems are different by their purposes and must be operated separately.
CAP control and management records provide data about farm structures and animal numbers, as well as some economic information such as the amount of the premiums. These data are stored in the statistics data bank along with survey data, which makes them available to all authorized users.
Surveys remain the backbone of statistical information. But, as surveys are usually neither exhaustive nor carried out every year, administrative data will prove very useful to supplement survey data with precisely localized data or to make for the absence of data between two surveys

Keywords : Common Agricultural Policy, Statistical Methods, Multi-Functionality, Administrative Statistics, Representativeness, Information Systems

1. Survey statistics and administrative statistics

While statisticians emphasize the conditions under which the use of non-survey data can be acceptable for statistics, local decision-makers have not waited for the blessing of the statisticians to begin using administrative data instead of approved statistics.
This has been permitted by the fact that many CAP records can give information on the structures of the farms, and, particularly in the case of the arable land premium records, the last available data about the acreage of crops may be as good, if not better, as the equivalent data of the last Farm Structure Survey.
More generally, administrative data and statistical data are both information on the farming industry. They may look similar, for their difference does not lie in the variables that they provide (acreage of crops, number of animals ...), but in the purpose of the action which produces them.
Basically, the purposes of administrative data systems are administration and control, not information.
If the purpose is to inform, to monitor the changes, to forecast, and to assess the result of agricultural policies, we are instead in the field of statistics.
Usually, administrative data flow up from the individuals to the administration in the framework of a declaration or claim procedure, as a result of the procedure itself.
Conversely, the statistician has to go for statistics with traditional collection procedures.

The main point is that in both cases there is a structuring of the data to set up the systems. The system manages economic and juridical data if the target is administrative, and is oriented towards the information of the public and decision-makers, forecasting, reporting and assessment if the target is statistical.

It can be thought that a single system supplied with administrative data could also allow to produce statistics, but this remains theoretical. If a decision-maker can be satisfied with some management data for his day-to-day tasks, such data cannot match all the needs of the statisticians, and the statistics service should have access to the information in order to restructure it. Technically, this would be possible, but in fact this would imply interrupting an administrative process which admits no delay, and is practically impossible.

The other way round is more difficult because the aim of statistics is not to manage individual data, but to use them to produce global data (means, aggregates …). Moreover, statistical data are under confidentiality and may not be used for other purposes than statistical information.

Another point must be underlined. It is a necessity for the administrative action to stick to the current realities. As a consequence of the economic and political change, concepts and the corresponding information will change as a function of the aims and priorities of the administrative action. For instance, the number of applicants for a particular premium scheme may change from year to year as a function of the changes in the conditions of attribution and will not necessary reflect the real change of the number of concerned holdings.

Unlike the administrative action, the statistical action implies setting up the tools necessary to monitor the changes, which necessitates a stability of the nomenclatures and concepts through the time and the space. Being independent of the policy implementation schedule, statistics have a better mastering of time. Values are given at a specific and significant dates, and short term data on the economic situation are still dependent on surveys, while administrative data are available only at dates depending of the premium scheme management, in the case where they are linked to a premium.

Both information systems may interact, thanks to interpenetrating, formalized links, or simple comparison points.

There is no absolute truth in that domain, and the situation must be considered in function of national cultures, for the situation is different between the European countries.

A country with a strong culture of administrative data management is in a position to easily use administrative information for statistical purposes.

In many other cases, where there is no similar interpenetrating, it is better to go no further than links or comparison points.

Formalized links allow to compare both systems and to explain the differences, for instance variables which are identical in both actions. Thus, crop acreages in subsidy claim forms may be considered as almost identical to crop acreages in structure surveys. On the contrary, the concept of holder as defined in structure surveys is different of the concept of applicant in the subsidy scheme, and one must be cautious when making a link at that level.

It must be noticed here that, when statisticians use well-defined concepts, administrative action uses a simple criterion of eligibility to the premium, which is never so well-defined and rich by its content as a statistical concept.

When there is no possibility of link, the statistician must rely on comparison points. In that case, the possibility of making comparisons is not permanent.

The problem of finding links or comparison points is not new for it also arises when matching agricultural statistics with general statistics, for instance in the domain of employment : a person who is classified as a farm head in function of the definition of Farm Structure surveys does not necessarily claim to be a farmer in the Population census, and, as mentioned above, is not necessarily the person whose name appears in subsidy claim forms.

At the end, it appears that to make comparisons or joint use of the data, it could be necessary to collect, within the administrative action, data which are beyond the purpose of the action.

Even if the concept may be considered as identical in both systems, there remains the question of the field. For instance, there is a trend to excuse small farmers to give detailed information in the subsidy claim forms. Information on small farmers must thus be gathered by conventional surveys.

Moreover, CAP subsidies concern many farmers, but not all farmers. They do not concern some part-time farmers, farmers whose production is not subsidized (wine, fruit, vegetables, organic farming, soil-less animal production, horse breeding…) or farmers who for personal or incidental reasons do not fill in a claim for premiums.

Once this problem has been formulated, if we are in an environment where there is a strong culture of administrative management of the farming industry, good information on each individual is available. Now, the question is to know if the administrative system provides all the required information, and how to supplement and harmonize this information.

To reach these objectives, we have seen that the administrative system must go further than its original purpose.

This is easier in a country with a limited number of farmers, and more difficult when there is a large number of farmers. What should be the size and cost of the system that we should have to set up to get all the detailed information which is required by statistics ?

If we compare Farm Structure Surveys and CAP data files, we see that crop acreage figures match very well, animal numbers do not much so well, but we see that there is no information about the farmer's family. We do not even know who is the farm head in the case of corporate forms. Additional questions which are in the Structure Surveys could be added to the subsidy claim forms (age of holder, number, sex and age of children, possible agricultural training of the children…). But this would mean asking those questions to about 450,000 farmers instead of the 70,000 of the Structure Survey. In the current context of reduction of the citizen paperwork burden, this is simply unfeasible.

Decisions, however, may rely on both systems, but must allow for an understanding why the figures are different, while being compatible. This understanding of the discrepancy must not necessarily be permanent, and a good approximation of it at given moments may prove to be enough. Nevertheless, the statistician must be able to match both systems at some aggregation level.

2. Design of a system

Either both systems are merged in a global system which includes both purposes, or two systems coexist. In any country, the point is to decide if it is preferable to set up a global system, in which it is possible in the same operation to obtain the statistical concept and the administrative concept, or to rely on separate systems.
A new element in this debate is that, as a result of the development of paperwork required by the new Common Agricultural Policy (CAP), and particularly with the Integrated Administration and Control System, the local services of the ministry of agriculture gather now a large amount of information about farms.
The French case shows that, if it had been decided to set up a global system, it would have been necessary, considering the number of concerned holdings, to introduce into the administrative system numerous items to make the necessary links, and this option was rejected.
It could be considered useful to associate the public agricultural statistics service to the design of the subsidy claim forms, or any administrative forms used to create files, to collect variables making links with statistical files.
On second thought, this does not appear to be a good idea, for, in a country were there are many holdings, this would mean to bother all of them by asking them detailed supplementary information that could have been obtained from a sample of farms.
It must be added that not only would such a procedure increase the paperwork burden of the farmer but, moreover, it would increase the data capture and processing work and, worst of, all, it could delay the premium payments, which must be avoided by all means.

3. The French SISA system

3.1. From miscellaneous uses of administrative data to a system

The local economic services of the ministry of agriculture can extract data from the Integrated Administration and Control System. The only unquestionable use of these data is their use as descriptors of the activity they are related to (amount of the subsidies, number of requests …). They are often used, however, to summarize the economic situation or to set up performance indicators for the local services. The statisticians are seldom asked to give their opinion about the validity of these data.
The use of administrative data is not new to statisticians, and for a long time, administrative data have been used as a supplement to survey data or to design sampling frames. In some sectors within the scope of activities of the ministry of agriculture such as vineyards and wine-making, agri-food industry, inventory of town and city facilities, environment … administrative data are much in use.
Nevertheless, it appears that there is no consistency in these uses, and the ministry wanted to make administrative data available to all authorized users in a national data base.
Considering that making data available and protecting their confidentiality was in the competence domain of the statisticians, the statistics office of the ministry (SCEES) was given the task to set up a national administrative statistics information system (SISA).

The data have been stored in the data bank of the SCEES, *Datagreste*, along with statistical surveys , in order to be stable, easily accessible and easy to consult. This base has been set up at the national level but is accessible to all local levels. Paradoxically, the availability to all levels of the administration of the ministry of agriculture is a consequence of the centralization of the files, which implies the same contents and the same definition of the items. The system permits a local statistician to analyse information on land belonging to farms for which the headquarters are outside of his administrative area, which is not possible using local subsidy claim records.
It is independent of the administration and control of the payment of support. The data put into *Datagreste* acquire the legal status of statistical data and are under the confidentiality of the French Statistical Law.

3.2. The contents of the SISA system
Although all the files which contain information linked to the administrative management of the farming industry may constitute sources of information, the contents of SISA are now limited to CAP subsidy management data.
CAP data are available at two levels :
- before payment, when all claim forms have been captured, from the Integrated Administration and Control System in the local offices of the ministry;
- after payment at the national paying bodies, usually "Offices" (product specialised state agencies).

At the beginning of the SISA project, as the Integrated Administration and Control System was not fully operational yet, it was considered preferable to extract the data from the paying body files. This allowed important simplifications and budget savings, the only expense for the SCEES being the checking of the data, their introduction into Datagreste, and the documentation.
However, the function of the paying bodies being to manage and control the payments, and not to produce statistics, there were some delays in the data extraction and some misunderstandings. Moreover, many farm characteristics which could have been useful but were not necessary for the payment procedure were not transmitted to the Offices.
Now, the control system has been set up under the name of PACAGE. It contains the farm characteristics that justify the attribution of the premium, the amount paid, and many details of the management of the premium which are useless to the statistician (delays in filling the form, presumption of fraud …). It is now possible to extract the information needed by the statistician from the PACAGE records.
This would allow to put into the national data base information which is not available at the national level, such as suckler cow premium, to collect information which is not transmitted to the national paying body and above all, to speed the availability of the information.
Once the data are available, there is still the problem of their validation (lack of coverage, discrepancy of concepts ...). We have seen that the introduction of links in the form of supplementary information conflicted with the inescapable trend towards administrative simplifications.
In such circumstances, the statisticians must limit their assessment capacity to the assessment of the consistency between aggregated estimates from both system. They have no control of individual records. However, they have obtained the presence of the

SIRENE firm register identification number, which introduces the possibility of further matching with statistic files.
What remains is that it is impossible to make of the presence of the SIRENE number a condition of the payment.
CAP data consist mainly of information about the management and control of the subsidy payments. Only a small part of it (basically farm characteristics and amounts paid) is interesting for the statistician and has been collected.
These data are representative at a very detailed geographic level for they are exhaustive for the field they are related to. Arable land under the arable land premium scheme represents 98% of the arable land according to SCEES's estimates.
This allows the use of CAP data as breakdown keys for survey data which are representative only at region (NUTS 2) level and not at lower levels (NUTS 3 or municipalities).
They can thus give yearly variation indices between two surveys and the geographic distribution at the level of the municipality of information already well known from traditional surveys at higher levels.
With this aim in view, CAP data have not been sampled, and all individual records have been stored.
Nevertheless, a sampling among the first subsidy claim forms has been carried out since the beginning of the premium scheme to get early estimates of arable land areas in May.
Now, the data base stores farm descriptive data from the CAP premium claims.
For six premiums:
- arable land premiums (cereals, oilseeds, proteaginous crops);
- non-alimentary fallow;
- ewe premium;
- special premium for male bovine animals;
- compensatory payment for less-favoured areas;
- premium for extensive animal production;

a file per premium and per year has been put into Datagreste since the 1993 season.

3.3. Other files

Any administrative source may prove useful, provided that producing statistics from it is not too costly.
An important resource under study is the very extensive administrative information resulting from the new tracing systems, particularly the file of the Individual Identification of the animals, set up in 1999-2000.
This type of source is not new to French statisticians. For many years, local files held by livestock offices of the agriculture chambers have been regularly used by statistical services to produce monthly and annual statistics, make farmers' accounts and the adjustment of structure or livestock surveys.
However, the problem that has to be faced is different, for that source does not contain information on local livestock, not even information on animal farms, but millions of individual identification records of animals. The work must thus be carried out on an unusual scale, with data which are not organized to describe animal farm structure but to ensure individual traceability.
If we compare with arable land premium records, the difference is obvious. Arable land records concern farms and the information which is collected is very close to the

information collected in Farm Structure Surveys. Both sources may be connected and compared. With some precautions in terms of viability, data of the arable land premium records could be substituted to some headings of the Farm Structure Surveys. This is not the case when it comes to contemplate substituting individual identification records of animals to livestock surveys.

Moreover, although the services in charge of these records willingly cooperate with the statisticians, the statisticians have in mind that their workload, due to incessant crises in animal production (BSE, foot-and-mouth disease ...) does not allow them to devote sufficient time to the cooperation with statisticians. At any rate, the statistician must not jeopardize the fundamental objectives of the administrative action.

4. Conclusion

The administrative data systems and the statistical system must develop in an independent way.

The task of the statistics service is to extract from a large bulk of control and management data those data which will provide the public and the decision-makers with information about farm structures, animal numbers and some data which are specific to the premium management and useful to economic analyses (e.g. amount of the premiums).

Thanks to the *Datagreste* statistical bank, these data, under the conditions of statistical confidentiality, are available to all authorized users, and not only as before to the local offices which produced them.

Surveys remain the backbone of statistical information. But, as surveys are usually neither exhaustive nor carried out every year, administrative data will prove very useful to supplement survey data with precisely localized data or to make for the absence of data between two surveys.

Possibilities and Preconditions for Data Integration. Experience in the Area of Agricultural Statistics by One of the Members of the European Union

Esa Ikäheimo
Information Centre of the Ministry of Agriculture and Forestry
P.O. Box 310, Helsinky, Finland
e-mail: esa.ikaheimo@mmm.fi

1. Population statistics

Over the last thirty years there has been a big change in the collection of the basic data for statistics in Finland, and this has had a significant impact on the operations of Finland's statistical authorities, the costs of compiling statistics and the authorities' ability to provide a service. Previously most of the data were gathered directly from the data providers, i.e. households, agricultural holdings, businesses etc. This was expensive and imposed an unnecessary burden on data providers, since the same information was gathered several times depending on the use to be made of it.

Finland's Population Register Centre was established in 1969. It keeps a register with details of persons, buildings and residential accommodation. Before register data were used, Statistics Finland's population censuses were laborious and costly. Basic data from the population register were used for the first time in the 1970 census, and this process became easier from 1975 onwards when register data were transmitted to Statistics Finland on magnetic tape.

Personal identity numbers were adopted for compiling statistics in 1971. This made it possible to combine a number of registers and improved the use of statistics for various reconciling operations. The general use of registers advanced to the extent that 1980 was the last time data for the population census were gathered separately from the whole adult population. Since 1990 censuses have been carried out entirely on the basis of the registers without any questionnaire-based survey. Census costs have fallen to a fraction of what they had been. At present 95% of Statistics Finland's total annual increment in the data stock is derived from administrative data

2. Agricultural statistics

Agricultural statistics are used to show the position of agriculture both in society and in relation to other industries. They are needed for the planning, research and counselling that backs political decisions in the sector. In addition, statistics are needed for assessing the effects of agricultural policy on rural communities, on agriculture itself and on society as a whole. The more agriculture is directed by society, the more there is a need for statistics.

Up to 1988 Finland's agricultural statistics were gathered by means of questionnaires sent by post or interview-based surveys. When the 1990 farm census was being prepared, a new register system was developed at the same time to serve both

administration and statistics. For the farm census and the new register that was to be set up, laws were enacted and passed by Parliament in 1989. It was laid down for the 1990 farm census that general data on census units could also be used, besides the statistical purposes of the census, for setting up the agriculture department's farm register.
General data are regarded as being:

- data identifying the census unit, i.e. name, location, land registry number and farm identification number,
- data identifying the owner or holder of the census unit, i.e. name, address, profession, personal identity number and business or company details, together with information on the census unit's areas by type of land use.

In the Act on the administrative farm register (later the Rural Business Register) the following provisions were laid down on the use of data in the register:
Register data can be used for the administration and supervision of aid to rural businesses, for other preparatory work on decisions, for planning measures and monitoring their effects, for statistical purposes and for research. In addition, register data can be used for supervising regulations and provisions on veterinary medicine and for the administrative measures this requires.
After the new register was introduced (in 1991) it was possible to discontinue almost entirely the questionnaires sent separately to farmers for updating the register, since some 95% of them had dealings with the administration because of aid applications, agreed changes to production or other reasons and thereby got into the register. In this situation it was possible to compile the basic agricultural statistics – mainly structural statistics – little by little as overall statistics, which did away with the problems associated with sample surveys regarding, for example, the comprehensiveness of the statistics and regional classifications.
The data content of the register was very limited, however, and did not satisfy the needs of statistics. In order to supplement the data, various statistical surveys had to be carried out. The register served as the basic population, however, for these mainly sample-based surveys. The need for agricultural statistics grew when Finland became a member of the EU in 1995. The amount of data to be gathered from farmers also increased when the EU's common agricultural policy was put into effect. Finland's register of rural industries was revised in the light of these needs so that it also met the requirements of the integrated administrative and control system (IACS).
Since the data in the register are also intended to be used for statistics, increasing the data content improved the usefulness of the register for statistical purposes. Data on the area under cultivation and a considerable amount of structural data are taken directly from the register. For statistics on farm animals, on the other hand, it has been necessary to find specific solutions, because the time frames, definitions and classifications of the data on animals in aid applications do not always meet the needs of statistics.

3. Possibilities of joint use

The decision taken in Finland shows that the joint use of administrative and statistical data considerably reduces the burden on farmers of providing data. In 2000, for example, about two thirds of the data for the farm census were obtained from the

Register of Rural Industries, which constitutes, thanks to its comprehensiveness, a sound basis for producing statistics. For administrative requirements alone hundreds of reports are extracted from the register each year.

Plenty of data on agriculture are also collected from the Population Register Centre, the Tax Administration, the Forestry Research Institute and the National Land Survey. Nowadays the records in the registers are in electronic form. In many registers personal identity numbers have been recorded, and these can be used as links when combining data. The use of location data is common in many data stocks, and this improves the possibilities for combining data.

It should be mentioned here that in Finland all field parcels have been digitised and entered separately in a register of field parcels which is updated annually in connection with aid applications. The digitised register of field parcels, which contains a GIS code, opens up new possibilities, e.g. for remote-sensing applications.

Of particular interest is the possibility of combining registers of businesses of different kinds. For example combining data on the farms operating in a district and on the small businesses set up in conjunction with these farms or operating separately from them gives an overall picture which can be used to assess the total importance of agriculture and business for the district. Similarly there is an opportunity to investigate the effect of rural policy measures directed at businesses of this kind.

The generalisation of IT and the development of computer equipment with regard to capacity and software has created new possibilities for the integrated use of data stocks. New angles have also appeared for the collection and use of basic data. More than half of Finland's 80 000 or so farmers have microcomputers. Approximately one farm out of three has an Internet connection. This makes it possible to an ever increasing extent to transmit electronically the information held on farms, for example to the Rural Business Register or directly to the statistical authority. If the needs of statistics and the administration are also taken into account when tax and accounting systems are being set up for farms, farmers can transfer the data needed for administrative and statistical purposes electronically without using paper forms. Practical applications are already in existence for this, although introducing them more widely will take a few years yet.

Statistical authorities have to see to it that information of interest to farmers is easily available at a reasonable price or free of charge. For example, weekly or even daily statistics on producer prices are important for farmers. Grain is also already traded on the Internet. In Finland there are producers' market organisations in operation which have as their business idea the marketing of for example vegetables and potatoes. Each day the producers report the prices and quantities of the products sold to the organisation's information system, where they are directly available to the organisation's members. The statistical authority has the possibility of using these prices for statistical purposes.

4. Conditions

A separate farm register for statistical purposes has been set up in Finland; this is used only for statistics and is updated with the data supplied by farmers to the Register of Rural Industries. Other databases have also been built up purely for statistical purposes. The statistical system must be kept separate from administrative systems. This is necessary if the principles of statistical ethics are to be upheld. It is also essential for the farmer to know when supplying data whether they are to be used for statistical or administrative purposes or both.

A second crucial condition is a legal basis for the joint use of data. Finland's general Statistics Act lays down that when work starts on compiling statistics an investigation must first be made of whether existing data can be used for the purpose. If data are already in existence and available, it is not permissible to undertake any separate gathering of statistical information. This principle also applies in the collection of agricultural statistics. In order to promote joint use the statistical authority has a statutory power to make use of data held by other authorities in compiling statistics. On the basis of this power the authority that draws up agricultural statistics can combine for example population-register or taxation data with information from farmers in order to produce its statistics.

A third condition is close cooperation between the statistical authority and the holders of data that are suitable for statistical purposes. Cooperation can help to ensure that the statistical point of view is already taken into account at the data-collection stage. Questions of definition and classification are particularly important here. This normally means that the statistical authority must be able to be involved as early as the stage of planning data collection.

On the technical side, the crucial condition for combining registers is the unambiguous identification of the units in the registers. Individuals are identified in registers of personal data by means of the personal identity number. This number is unique and identifies an individual precisely. In business registers and farm registers identification is not so clear and simple. For example, a farm is normally made up of several different plots of land, each of which has its own identification number.

In Finland's Register of Rural Industries each farm has been given a unique registration number which is made up of the number of the municipality, the farm's serial number within the municipality and a check segment. However, this farm identification number is used only in the Rural Business Register, and the tax authorities and the land registry, for example, have their own identification numbers for farms.

Preparatory work is starting in Finland just now for integrating the farm register with the business information system. The aim of this project is to standardise the presentation of farms in administrative registers and at the same time make data collection and the joint use of registers more effective. The project will trace the relationships between the administrative registers used by the keepers of various registers to delineate farms, analysing the identification numbers and concepts used and endeavouring to further promote the integrated use of registers.

Another important precondition for the joint use of registers is the development of IT facilities. The registers must be set up systematically and attention must already be given at the planning stage to the possibilities of joint use.

The joint use of registers has confronted the compilers of statistics with new demands. In addition to the traditional collection of statistical data they have to master the necessary information technology for exploiting the stocks of data in various information systems. This means a change in statistical work, in that instead of the traditional questionnaire-based collection of data a command is needed of non-statistical information systems and of the use of computer software and equipment, and a corresponding change is necessary in staff structures.

5. Conclusions

A characteristic feature of the production of statistics is continuity, which means long time series and the comparability of statistics over time. The methods of compiling statistics, however, have changed irreversibly, from the traditional gathering of data to the use of existing data stocks. This development is to be welcomed for the opportunities it offers and the conditions for it need to be consolidated.

The statistical authorities are facing constant pressure to cut the costs of producing statistics while at the same time being expected to improve statistical services. Better services mean more extensive, quicker and more analytical information than before on agriculture and other related industries and on the environmental effects of agriculture. At the same time farmers – the providers of statistical data – are suffering in the grip of the red tape caused by EU bureaucracy. A significant increase in statistical surveys of agriculture is not possible, nor would it even be sensible.

In providing services to the users of statistics the statistical authorities must shift the emphasis of their operations and their use of resources from the gathering of data to the recasting, processing and analysing of existing information and to user-friendly dissemination. Suppliers of data can also be motivated by an improved information service, which is made possible by electronic data links.

Promoting the use of existing data for statistical purposes is particularly worthwhile from the cost viewpoint. The full cost savings in agricultural statistics in Finland have not been felt immediately, since switching to a new statistical culture requires capital expenditure, both on IT systems and on human resources. In addition it is also necessary in the transitional phase to maintain the old statistical system, so that comparisons can be made to ensure the quality of the statistics produced using the new system. Quality assurance has come to be particularly important where existing data sources are used. In the long term cost savings can nonetheless be expected. The direct benefit lies mainly in reducing the burden of supplying data and in improving the service provided.

In Finland's experience the most crucial preconditions for the joint use of data stocks are a legal basis, common rules of the game for "good register-keeping practice" by the various keepers of registers, the compatibility of information systems and smooth cooperation between the keepers of registers and the statistical authorities. Even close cooperation must not, however, mean placing statistical and administrative data in the same database. From the point of view of statistical ethics it is an essential condition that the statistical system is kept independent and clearly separated from the repositories of administrative data.

ESSAYS

n. 12/2002

Conference on Agricultural and Environmental Statistical Applications in Rome

Rome, 4-8 June 2001

Volume III

ISTAT

7 June 2001

Plenary Session III

TECHNOLOGICAL IMPACT ON DATA MANAGEMENT

Chair: A. Lindner

The Data Warehouse, a Modern System for the Dissemination of Information

Enrico Giovannini[1], Alberto Sorce[2]
National Institute of Statistics– ISTAT (Italy)

Abstract: remarkable modifications in the organisation structure of national statistical institutes have been adopted to solve various problems; such as the adoption of new European regulations, the growth of users requirements, the reduction of statistical burden on respondents, and finally continuous innovation in the field of computer technologies. Thanks to IT developments during recent years, it is now possible to develop a unique environment of integration for statistical data, linking both the processes and the databases locally developed. In particular, the development of Data Warehouses permits the integration of data coming from different surveys and the dissemination of large databases, which can fulfil different users' needs.

Keywords: integration, dissemination, friendly tools, economic statistics, and information statistical systems.

1. Introduction

The organisation and the activities of National Statistical Institutes in Europe are rapidly changing for different reasons. Some of the main factors are:

- the adoption of new European regulations in the field of statistics;
- the growth and diversification of the users' requirements for statistical information;
- the necessity to further reduce the statistical burden on respondents;
- the continuous innovation in the field of computer technologies.

These reasons imply remarkable modifications in the organisational structure of National Statistical Institutes (NSIs) and in the statistical methodologies adopted.

This paper deals with the development of the Statistical Information System on Enterprises and Institutions (SISSIEI), designed and developed by the Department for Economic Statistics (DISE) of the Italian Statistical Institute (Istat). SISSIEI represents a multidimensional structure devoted to creating an organic picture of all the statistics produced by the Department with reference to agricultural holdings, enterprises, and public and private institutions. The system has been drawn up as an element unifying the statistical activities of the enterprises and the institutions, for the purpose of producing statistical information more efficiently, to increase the quality of data, and to rationalize internal information flows. It implies a real cultural "change" in the way surveys are conducted: in fact, as with the system of national accounts, SISSIEI intends

---

[1] Chief Statistician of OECD (enrico.giovannini@oecd.org)
[2] Responsible of statistical information system on enterprises and institutions of Istat (sorce@istat.it)

to represent a structure where all single statistical processes are represented, encoded and integrated.
SISSIEI covers all phases of statistical data collection, processing and dissemination in agricultural, industrial and service sectors. In particular, this paper is focused on dissemination aspects, in order to evaluate how the so-called “data-warehousing” approach can help an NSI to disseminate large amounts of statistical information using new Internet technologies.

2. European Union regulations and the organisation of National Statistical Institutes

The European Community regulations emanated in the last few years in the statistical field, in particular in economic statistics, can be considered single "bricks" which can be used to create a single more complicated “information statistical system”. The regulations concerning statistical units, classifications of economic activities, registers of enterprises, structural and short-term economic statistics, national accounts and specific sector statistics (agriculture, tourism, transports, etc.) have been based (in general) on coherent concepts, definitions and classifications. In this way, EU institutions have obliged member states to implement notable changes in the organisation and methodologies used to compile different statistics. At the same time, national statistical systems have had to pursue a more elevated level of efficiency, managing situations in which they simultaneously experienced some reductions in their resources in real terms, a large increase of users’ needs and strong pressure to reduce the statistical burden on respondents.
In Italy, the implementation of these new statistical regulations has been interpreted as an opportunity to develop a new system of economic statistics, by putting together the organisational, methodological, institutional and technological aspects, in order to:

- extend the use of administrative sources for statistical purposes;
- develop integrated systems of metadata;
- improve the co-ordination of statistical methodologies adopted in different surveys;
- implement new IT instruments and adopt new information systems, moving from the mainframe technology to distributed systems;
- involve users in the definition of new products, and respondents in the implementation of new data collection systems.

The development of a systemic design brought about in 1997, the reorganisation of the Department of Economic Statistics (at that time called "Central Direction for Statistics on Enterprises and Institutions"). This reorganisation was based on the creation of Divisions which have to manage “similar” production processes independently from the economic sector. The adoption of a "process based" approach has been completed with the recent creation, within this Department, of the Central Directorate for Structural Statistics, the Central Directorate for Short-term Statistics, and the Central Directorate for Economic Censuses and Registers. This approach led to a large improvement in the

efficiency of statistical processes and in the coherence of methodologies used for similar surveys.
This tendency to reorganise statistical activities "by process" is common to other European NSIs. This has helped international relationships and co-ordination among homogeneous structures inside different member states of the Union, among which the exchange of experiences has notably grown.
Of course, this kind of organisation could create some risks in terms of coherence of output. If statistics referring to a certain sector (for instance, industry) are carried out in different Divisions, there is a potential risk of losing the coherence of the information disseminated. From another point of view, it is more difficult to develop some integrated analyses of that sector, because the sector experts are spread over different Divisions. As the users' needs are mainly focused on integrated presentations of statistical data for the same sector (or subject), it is essential that a single productive process be seen as part of a larger statistical system. This implies that organisation "by process" should be accompanied by the development of tools that allow the integrated presentation and dissemination of available data.
The construction of SISSIEI responds exactly to these requirements, assuring interdependence among the structures responsible for the production of different data, allowing a full integration of the data production, up to the level of microdata. At the base of SISSIEI there is a detailed "vision" of the productive process, based on the concept of *corporate data warehouse*, whose the main characteristics are described in the following paragraph.

3. SISSIEI: a system based on the "corporate data warehouse" vision

Traditionally, the activities of NSIs have been directed towards the management of the surveys, and the production of the relative results. It is evident that, with the growth in demand for information and, above all, with the increasing need of the user to have a complex informative frame, the control and the coordination of different (and linked) surveys become very difficult. A possible solution to this problem is the development of modern statistical information systems.
A statistical information system is typically composed of subsystems that allow the collection, the processing, the storage, the analysis and the dissemination of statistical data (Giovannini, E., and Egidi, V., Istat, 1999). The statistical process uses and produces statistical data whose elementary components are the microdata, the macrodata and the metadata. Obviously, each statistical organisation can give different emphasis to certain types of statistical processes, and has different ways of defining the microdata or the macrodata[3].
A statistical information system, after having been designed to produce specific statistical products, could have further general purposes, for example, to provide users with "statistical services" (i.e., the extraction of samples from statistical registers for "private" or "public" surveys). In this sense, the information system becomes a common infrastructure, in which different types of subsystems develop a specific task and interact with each other.

[3] For instance, macrodata referred to the population of people and of enterprises, collected by a National Statistical Institute, can be considered by international organisations as microdata of a supra-national statistical information system.

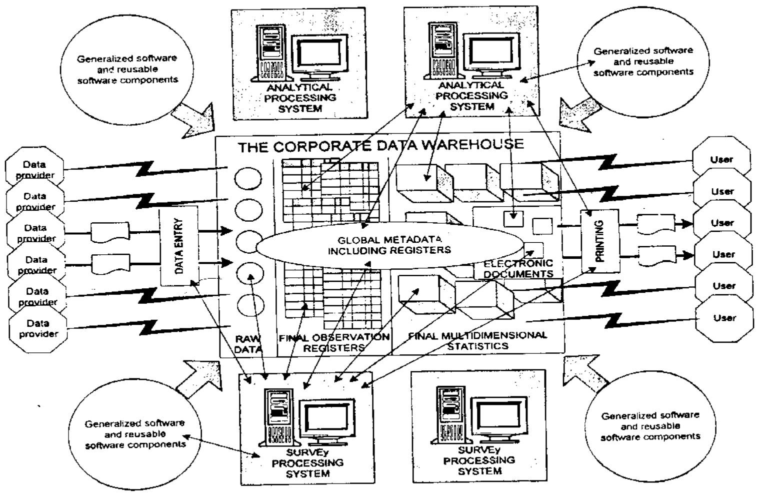
Due to recent technological developments, it is now possible to develop a unique environment of integration, linking both the process and the locally developed databases. Accordingly, the function of integration has to be very flexible and allow direct usage by different users, through some *user-friendly* tools. On the other hand, in order to allow authorised users to fully navigate the data deriving from different statistical surveys, the function of integration has to be based on a very well designed system of metadata.

A possible "vision" of an integrated information system is contained in Chart 1, created by the U.N. (1999), based on the development of a "corporate data warehouse". The system is articulated in different areas (data collection, data processing, data dissemination), and different productive divisions (analytical processing systems) can interact among them through generalised software. At the end of the process, the final data are stored in the "final observation register" area, from which multidimensional objects (so-called "hypercubes") are used to create final printed and electronic products.

To manage this system a "centralised" unit is absolutely necessary. This centralised function of integration has been developed (in the case of Istat) at departmental level. Obviously, good cooperation among the centralised and decentralised structures should exist, but to ensure efficiency of the system, it is necessary to define the responsibilities of different structures very clearly, in order to avoid redundancies and duplications.

The centralised unit has to facilitate the data exchange (as well as the metadata) among the structures. In a sense, it operates as the "pivot" of a very complex network, where statistical and technological competencies are working together, using different languages and having different needs. In many cases, this approach imposes a cultural "revolution", to move from a fragmented organisation, in which the autonomy of each single unit is substantially safeguarded, to an integrated organisation, in which each unit depends on the others.

Chart 1: *An architecture of informative system for statistical organisations*



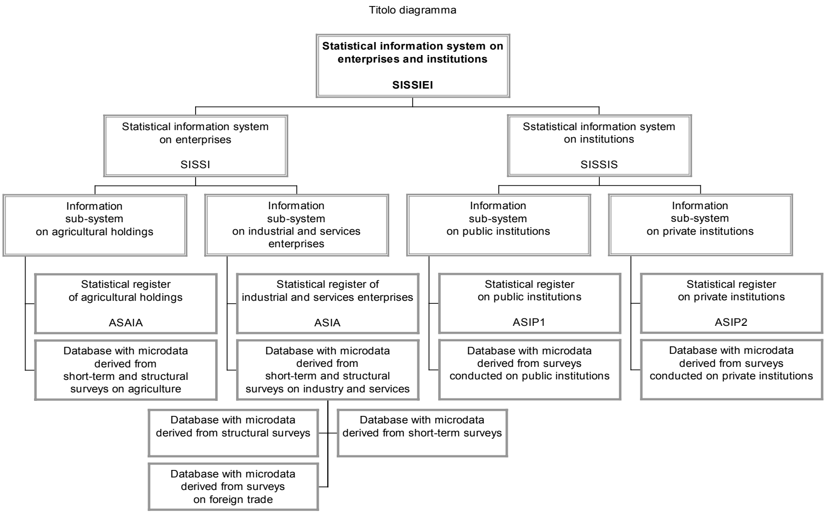
The Statistical Information System on Enterprises and Institutions (SISSIEI) fully adopts this vision (see Giovannini and Sorce, 2000; Calzaroni, Giovannini and Sorce, 2000). In particular, SISSIEI is based on different navigators to explore databases of statistical data and the related data dictionary which allows the connection of variables derived from different statistical surveys. The System permits *on-line* access to microdata by employing SQL commands. For instance, through an OLAP process (*On-line Analytical Process*), it is possible to get on single display all the microdata derived from different surveys and referring to a single enterprise, integrated with the information extracted from the statistical register of active enterprises (address, juridical form, sector of activity, number of employees, etc.). In addition, SISSIEI offers internal users some generalised tools for extracting lists of units (enterprises, agricultural holdings, etc.), and for definiting samples, etc.
Chart 2 presents the main areas in which the System is structured. It is articulated in two areas: the first refers to agricultural holdings and to industrial and service enterprises; the second to public and private institutions. Each subsystem is based on the relative register of economic units, according to the European Community regulation n. 2186/93. In particular, the enterprises' subsystem utilises the statistical register of agricultural holdings (ASAIA, with about 2.800.000 units) and the register of industrial and services enterprises (ASIA, with about 3.500.000 units). In the case of institutions, the statistical register of public institutions (ASIP1, about 10.000 units) is complemented with the register of private institutions (ASIP2, about 200.000 units).

4. The data warehouse for dissemination purposes: some examples

When the final data have become available, SISSIEI serves as a base for the construction of data warehouses (DW) for the dissemination process. In particular, three DW have been created between 1998 and 2000 and Istat is now developing two new DW for the 2000 agricultural census and the 1999 census of non-profit institutions. In particular, in 1998 Istat produced and made available on Internet its first DW for the dissemination of the 1996 intermediate census on industrial and services enterprises results. The success of this new tool has been widely recognised by users: the database was made freely available on Internet at the end of 1998 *(http://cens.istat.it)* and during 1999 more than 2.500 users have made more than 300.000 extractions of statistical tables. Moreover, from the same database Istat has automatically produced 20 CD-ROMs and 120 paper publications, where results disaggregated at province and municipality levels were presented.
The planning and realisation of the database involved more than six months of work. The DW was implemented using SAS/Warehouse Administrator (for the construction of the Date Warehouse) and SAS/Desktop Application Builder (for the choice of aggregations and the construction of the application for navigation). Users have the possibility of extracting tables (with different formats) from about one million pre-calculated cells: the dimension of the DW is about 25 GBs. It contains not only data collected in the 1996 intermediate census, but also those derived from 1971, 1981 and 1991 general censuses on industry and services[4].

[4] In particular, while the 1971, 1981 and 1991 data (only referring to enterprises) were stored in three different files, one for each year, the 1996 data were stored in two files, one for enterprises and one for

Chart 2: *The general structure of SISSIEI*



During the year 2000 two new DWs were developed, and placement on the Internet is foreseen for this year: the first refers to the results of the annual structural surveys on industrial and service enterprises; the second contains foreign trade statistics.
The DW on structural business statistics was designed to contain the results of surveys conducted on economic accounts of small, medium and large enterprises, as well as the results of the second phase of the intermediate census on industries and services. It will be periodically updated and enlarged with data derived from all annual business surveys, i.e. on the structure of labour costs, on technological innovation, on R&D, etc.
The complexity of the project is due to the fact that the variables to be published are derived from different surveys, with different keys of aggregation, like territory, size classes, sectors of economic activity, etc. Moreover, these data can be disseminated only after a careful check to avoid any violation of confidentiality.
The DW containing foreign trade data is different from those mentioned above because it is fully inserted in the production process of statistical data. In fact, it was initially built for the realisation of around 1.000 tables contained in the statistical yearbook "Foreign Trade and International Activity of the Enterprises", published for the first time in 1999 by Istat and the Italian Institute of the Foreign Trade (ICE). Subsequently, the project has been widened, to integrate the DW inside the production process for foreign trade monthly data and to offer external users rapid and efficient access to

local units. To ensure a full comparability of data derived from different censuses, some statistical operations have been carried out, like the harmonisation:
- of the classifications of economic activities;
- of the administrative limits for municipalities and provinces;
- of the classification by juridical form characteristics;
- of the code for the artisan nature of each enterprise whose definition varied across different censuses.

aggregated figures. In particular, from the beginning of 2001, the Istat regional offices have access through the Intranet to:

- "static" tables available on the Web Server, including those contained in the on-line version of the above mentioned statistical yearbook;
- the *data mart* available on the *Application Server*, in order to extract tables on aggregated data (according to default formats);
- the *facts table* and to microdata memorised in the DBMS Oracle for more detailed applications.

Currently, the DW contains monthly microdata referring to the period 1991-2000 for over 9.000 groups of products, disaggregated for 250 countries, 103 provinces of origin or destination of goods. The volume of data is quantified in around 80 Gbs. This dimension characterises the DW as a Very Large Data Base (VLDB). From this DW, a more limited database (with only aggregated data) will be made accessible to external users on the Internet.
The technology used for the DW on foreign trade is Microsoft (Asp technology), while data are stored in an Oracle DMBS. The technology used for the DW on structural business statistics is different and based on an Oracle DBMS, but the creation of the DW and the dissemination phase use Business Objects. This change is due to the experience in 2001 with the fourth DW, containing data obtained by the first general census conducted on 400.000 private (no-profit) institutions.
Also in this case, the DW is fully integrated with the production process. In particular, in the data collection phase, Visual Basic was used to allow 103 Chambers of Commerce to perform, via Internet, the data entry phase, while in the rest of the process three different technologies were used: SAS for the creation of facts tables; Oracle as a repository for the memorisation of the "star scheme" and of fact tables, as well as DBMS; Business Objects for the creation of tables for the data check, and the creation of the final lists of institutions, for the publication of statistical tables and for access to final results through Internet.

Table 1: *Products used for the development of Data Warehouses*

Subject	DW Administrator	Repository
Intermediate census on enterprises	SAS	SAS
Foreign trade	Microsoft	Oracle
Structural business statistics	Webi/Business Objects	Oracle
Census on private (no-profit) institutions	Webi/Business Objects	Oracle

After these experiences (summarised in Table 1), Istat has not chosen a standard technology for the construction of its future Data Warehouses. This uncertainty is mainly due to the fact that the technology is evolving very fast and it is necessary to follow continuously the new tendencies in order to maximise the quality of the results. A keyword in this area is the "partnership" between the user (in this case the NSI) and the vendor of software; a second keyword is "integration" between different technologies, in order to use the best software for each phase of the process. Finally, the success in this field can only come from a close co-operation between statisticians and IT experts and it is absolutely necessary to have some of these experts working inside the NSI, in order to maintain the control of the relationships with the vendors of software.

4. A new data warehouse for the dissemination of the agricultural census results

In 2000 Istat carried out a general census on agriculture. Field operations were conducted between October 2000 and January 2001, with about 20.000 enumerators and co-ordinators. The census was monitored through the Internet: in particular, the Chamber of Commerce, the statistical offices of the Regions, the Istat regional and central offices had the opportunity to continuously follow the evolution of the census, looking at the reports prepared by the co-ordinators operating at the municipality level. The data contained in these reports were imputed directly on to the Internet by the co-ordinators, or were transmitted by e-mail (or fax) to the Chamber of Commerce, which imputed the data in the monitoring system. The registered data are currently in the quality check phase: the preliminary results will be published at the end of June, while the final results will be disseminated in the second half of 2001.
For the dissemination of the final results a DW is under construction. In particular it will have the purpose of publishing the statistical data derived from the last five general agriculture census, making them available on the Internet through interactive and personalized applications. For such a purpose it has been decided to develop the DW with the technology Business Objects / Webi, while the data will be stored in an Oracle DBMS.
The project foresees the creation of a data mart inherent in the various typologies of agricultural activity. The areas for which data will be presented are the following:

- General information about agricultural holdings,
- Use of the ground and production,
- Irrigations, rural buildings and plants,
- Information on livestock,
- Use of the agricultural machines,
- Employment,
- Other information (use of information technologies, environmental aspects, etc.).

Several dimensions and classifications can be used to present census data. Some limitations will be imposed to avoid the disclosure of personal data, taking into account the fact that the law authorises Istat to disseminate the information necessary to fulfil supranational, national, regional and local needs. The results will be disseminated for 8.103 municipalities, 103 provinces and 21 regions: in addition, the results will be integrated in a new Geographical Information System with the data which will be obtained from the 2001 general population and economic censuses. Finally, Istat is studying the possibility of publishing the agricultural census data following new classifications, taking into account the structural characteristics of the holdings. Some experiments conducted on the 1990 census data were considered very useful and interesting by the researchers, mainly in service enterprises. The success of this new tool has been widely recognised by the users, mainly in understanding the characteristics and the development of some new typologies of agricultural holdings.

References

Calzaroni M., Giovannini E., Sorce A. (2000) "Il Sistema Informativo Statistico Sulle Imprese dell'Istat: problematiche e potenzialità", presented at the XL Conference of the Italian Statistical Society, Florence, 26-28 April.

Giovannini E. e Egidi V. (1999) "Sistemi informativi integrati per l'analisi di fenomeni complessi e multidimensionali", *Atti della Quarta conferenza nazionale di statistica*, Istat, Roma.

Giovannini E. e Sorce A. (2000) "The Statistical Information System on Enterprises and Institutions of Italian Statistical Institute: a Datawarehouse Approach", presented at the 9th CEIES Seminar on "Innovations in Provision and Production of Statistics: the Importance of New Technologies", Helsinki, 20-21 January.

Giovannini E. e Sorce A. (2001) "The impact on data warehousing on the management of statistical offices" presented at the "Meeting on the Management of Statistical Information Technology 2001", Geneva, 14-16 February.

Information Management and Dissemination: Offering Access to a Comprehensive Cross-Section of Statistical Data and Metadata in Support of Multidisciplinary Projects

Peter Lübkert
Head of Systems Development and Support
OECD, EXD/ITN[1], Paris, France
e-mail: peter.lubkert@oecd.org

Abstract: The rapid and easy access to statistical information is a major challenge for many organisations. This paper focuses on work carried out by the OECD Secretariat to render its statistical data and metadata accessible for different constituencies. It gives a brief overview of today's analytical/statistical computing and communication requirements. A key success factor for the horizontal integration of different sets of statistical data and metadata is the availability of a central catalogue, which contains comprehensive information about the various data sources. The concepts outlined in this paper have been developed in close collaboration with the Statistics Directorate and through a specialised internal task force[2] which includes major users of statistical information at the OECD.

Keywords: statistical information management, statistical data-/metadata repositories, data-warehouse, information dissemination, statistical portal services.

1. Introduction

The OECD is recognised throughout the world for its reliable and comprehensive analytical and statistical work. The gathering and harmonisation of international data in a multidisciplinary environment are key to international comparison and policy making. Statistical, analytical and policy work is essential to meet the current as well as medium and long-term business challenges of the OECD. It enables OECD committees to carry out a wide range of assignments in today's global economic and social landscape. International trade, economic forecasting and sustainable development are only a few examples of issues which heavily rely on this area of activity, and which illustrate the diversity of themes that need to be addressed.
Statistical computing activities at the OECD encompass the collection of national statistics, their validation and harmonisation, the derivation of new internationally-comparable indicators, economic policy analysis, data modelling, and publishing in both paper and electronic forms. To carry out this work, OECD economists, analysts, and statisticians require a comprehensive set of tools ranging from the standard spreadsheet

[1] The OECD Executive Directorate's Information Technology and Network Service
[2] The OECD Analytical Statistical Task Force (ASTF) - see also acknowledgements

package over more specialised data manipulation and management systems to a variety of econometric and modelling software packages.
The gradual migration of mainframe-based legacy applications and systems towards a more modern computing platforms in the 90's has allowed us to keep up with evolving user needs, to adopt emerging statistical standards (e.g., SNA93), and help to absorb increased pressures on resources. Following the completion of this migration, development work is now focusing on improving data collection, streamlining statistical processes, facilitating horizontal work, and shortening production cycles.

2. The Context

The OECD is a knowledge-based organisation with a multi-faceted working environment and a wide spectrum of information management and dissemination needs. Many of its activities heavily rely on the availability of specialised ICT tools, which need to communicate well among each other.

2.1 Management of Corporate Information

The Organisation creates and requires access to a substantial amount of information. Approximately 12,000 official documents and 500 publications are issued each year, in addition to the regular dissemination of statistical databases. Much progress has been made in introducing common facilities for the authoring, exchange, translation and dissemination of OECD information. Common corporate repositories are increasingly being used to store final information, for internal and external access. Managing these information bases is a major task which requires a sound IT infrastructure.
All OECD information is generated, stored, issued and re-used in electronic form. The widespread acceptance of the Internet has opened a new dimension for rapid and simple dissemination of and access to information. The continuously evolving international data communications infrastructure combined with Internet-based information dissemination technologies will not only help to shorten data collection and publication cycles, but also to address other business priorities such as the management of corporate information flows and outreach activities. The implementation of common functions to integrate the collection and initial storage of data and metadata will contribute to improved timeliness, elimination of errors and duplication, and reduced reporting burden on Member countries.

2.2 OECD Communication Strategy

The OECD communications strategy is an important element in ensuring an effective and coherent interaction with Member countries and the outside world in general. This interaction is facilitated through a range of technologies: from traditional telephone, e-mail, fax, and online information services (OLIS), to more recent developments such as Internet, OLISnet, committee discussion groups and videoconferencing. The OECD Statistics Directorate, as well as other departments involved in analytical and statistical information processing make considerable use of modern communications technology, including committee discussion groups (e.g., the Discussion Forum of the Task Force on International Trade Statistics).
OECD information is available to Member countries 24 hours a day, 7 days a week. The OECD's international communications network provides fast, reliable and secure

information exchange and dialogue between Member countries and the Secretariat through OLISnet services. One-stop access to a wide range of OECD online information and services is provided to a privileged audience of over seven thousand Member country officials in Delegations and approximately four hundred ministries, statistical offices, central banks, and institutions. Online information available includes official documents, publications, statistics, e-mail, and committee discussion groups. In addition, OLISnet services provide access to the OECD Internet site and selected information from the OECD Intranet site.

The OECD Internet site itself provides a range of public information about the Organisation’s activities, as well as committee discussion group facilities for OECD correspondents in academia and research institutes.

Electronic Data Products play an important role in disseminating OECD information to Member countries, international organisations, and the private sector. In collaboration with Public Affairs and Communications Directorates, the ICT[3] Service (ITN) is developing uniform software tools to be used with all OECD data products, regardless of the media used (CD-ROM, on-line, diskette, etc.). Products with a fixed update frequency are complemented with "real-time" data sets kept up-to-date in the internal Corporate Data Environment and made available selectively to clients via the Internet, as well as OLISnet. The expansion of the range of statistical information available on the Internet will bring OECD statistics to the client base in existing as well as new formats. Further developments of the Electronic Data Products programme will focus on providing one stop access, where authorised, to a wide range of online international comparable statistics through OLISnet and the OECD Internet site (i.e. statistical portal services).

3. Requirements

Collections of statistical information are used internally for analytical work, econometric modelling and forecasting, and subsequently in publications. In addition, the online access to published OECD statistical information has become a priority in the past few years. Offering easy online access to OECD statistics is one of the Organisation’s strategic management and communications objectives, as well as a medium-term information technology direction.

The OECD computing environment has to integrate a variety of software tools, and provide for easy sharing of information among applications residing on PCs, workstations, and networked servers. In this context of particular importance is the availability of a user-friendly data management software environment, unifying the Organisation's various databases and meeting the diverse requirements of both database managers and end-users. The increasing importance and availability of metadata compound this requirement. New methods for the collection of statistical data must be implemented to accommodate evolving international reporting standards and take advantage of new transmission media.

Equally important is the availability of analytical software for econometrics, statistical analysis, macro-economic modelling, and graphics which, in turn, integrate with

[3] Information and Communications Technology

standardised procedures for the preparation of compound documents (integrating text, tables and graphics) for internal and committee use, and for external distribution.

Last, but not least it is essential to constantly increase the speed, quality and flexibility of the Organisation's publication production and printing processes and to enhance facilities and procedures used for electronic data capture and data dissemination.

The rapid evolution of ICT technologies make it feasible to envisage innovative, cost-effective solutions to these computing needs, while at the same time providing better access to user-friendly, commercial software tools and database management and manipulation facilities. Also, the Analytical/Statistical systems architecture must evolve in harmony with overall ICT developments and leverage existing investments.

Opportunities exist for increasing the efficiency of information production and management, and for improving the quality and effectiveness of presentation to targeted audiences, both internal and external. Requirements for offering access to OECD statistical information are as follows:

3.1 From the users' point of view

- More integrated analytical/statistical working environment
- Easier to use data manipulation software and easier access to data
- Simpler data collection and improved metadata handling
- Availability of modern analytical tools/packages
- Easier, better automated, higher-quality graphics production
- Extended interactive graphical data analysis facilities
- Shorter publishing cycles
- More attractive work (IT) environment for outside professionals (new recruits, consultants)
- Flexibility to meet future end user needs, shifts in technology

3.2 From ICT service point of view

- Opportunities to further standardise and streamline the overall ICT infrastructure
- Adoption of emerging industry standards and international norms
- Increased application scalability
- Portability and inter-operability of analytical and statistical information systems across a range of different operating systems and manufacturers' equipment and a range of equipment of different computing capacity.

3.3 Data

There exist two main types of data objects to be stored and managed at the OECD in support of analytical/statistical applications. The most common type is time series, i.e. a vector of numeric observations ordered by time, associated with various textual attributes such as the time series name and description (e.g. agricultural production for Switzerland, total at current prices, in million francs), the frequency of observations (e.g. monthly), starting period, etc. Today time series objects are generally managed centrally in a common SQL-based systems environment and with FAME[4], a time-series manipulation and graphics tool.

At the OECD, there also exist many sets of data values, which are homogeneous enough to be structured as multidimensional data arrays. Typically, one dimension of such

[4] Forecasting, Analysis, and Modelling Environment

arrays is time and another one country (i.e. the list of OECD Member and partner countries), while other variable dimensions are more specific to the application. For example, for foreign trade data by commodities, other dimensions include the list of trading partner countries, a normalised list of products (SITC, HS), value or quantity, and the type of flow (imports, exports). A cell of the array, i.e. an observation, then corresponds to a unique combination of one element in each dimension. One obvious advantage of multidimensional arrays over individual time series is that data indexing comes as a natural by-product of the multidimensional structure definitions. Another advantage is the ability to process at once logical groups of many time-series and, thus, simplify group operations such as data retrieval or aggregation. However, when time series data sets are volatile or heterogeneous, it may prove more efficient to treat time series individually. Commercial software offerings using the concept of multidimensional data objects became available in the 1990's. At that time the OECD chose IRI Express (today ORACLE Express) as their standard product for the manipulation of this kind of data.

3.4 Metadata

Information about data objects, i.e. metadata, also needs to be maintained along with data objects. Reference item lists, such as nomenclatures and/or dimensions of multidimensional data arrays (e.g. list of OECD Member countries) are obvious examples of metadata objects. These lists often need to be hierarchical (e.g., countries within geographic zones) and sharable amongst database users, so as to facilitate structure harmonisation across databases. Metadata also include descriptive information, documentation and notes at different levels (e.g. at the application database, time series or even individual observation level) to describe important characteristics of data object contents. In the context of application development, metadata are a challenge and tool at the same time, meaning that metadata always has to "travel" with (be attached to) the data, but also has to help locating the data.

3.5 Catalogues

The development of a central catalogue of all officially available application databases and their respective contents is very important. It should include data sources, committees, user directorates, etc. - and about application databases - including associated data collection methods, database system implementation, database usage, access control rights, database output products (publications, electronic data products), etc.

Obviously, not all application data objects used need to appear in the catalogue. A distinction thus needs to be made between catalogued application databases (i.e. clean, official, shared groups of data objects), on one hand and non-catalogued databases (i.e. unofficial working data-sets) on the other. For instance, incoming data from Member countries will, in general, transit first through non-catalogued databases, to be validated and checked for consistency, before becoming catalogued and thus part of the central data environment.

3.6 User Communities

When deciding on how best to provide access to a cross-section of statistical data and metadata in support of multidisciplinary projects, it is particularly important to differentiate user communities. Two major categories can be distinguished - internal and external users. In a recent study on improving access to OECD statistics for internal users at the OECD were categorised by:

- statistical assistant;
- database manager;
- economist;
- visiting expert;
- management;
- IT specialist.

The study highlighted the importance of the following needs for internal users of OECD statistics:

- locating OECD 'Reference' series (most popular series - i.e. Exchange rates, CPI, GDP, Unemployment);
- locating data by theme from different sources;
- global keyword search across directorate for specific series or across an entire catalogue;
- making your data more visible/accessible to others;
- clarifying/highlighting areas of duplication;
- monitoring data access.

Some of the key needs identified for internal target audiences extend clearly also to external users (notably the first three in the list above) and have to become part of related data dissemination strategies.

4. Putting IT Together

The successful implementation of a strategy to develop and keep up-to-date an ICT infrastructure in support of analytical /statistical processing is a complex and multi-faceted challenge. Obviously information and user requirements come first, but elements like ICT strategy and a management structure are also important.

4.1 Governance

A proper management framework with executive-level support is critical for any large ICT investment. Objectives, costs and expected benefits have to be spelled out from the beginning. A clear vision statement underpinning the business relevance is essential. Further technical directions have to be adopted and co-ordinated. The OECD Statistical Policy Group (SPG) is a senior management group chaired by the Chief Statistician[5], which takes the lead in the co-ordination of organisation-wide statistical strategies. A Board of Directors for Computer and Communications Strategies provides global guidance for ICT investments and agrees on overall strategic directions. An inter-directorate Task Force on new technologies for Analytical/ Statistical systems (ASTF) carries out the more detailed analysis of technical strategies and implementation plans. The ASTF co-ordinates its efforts very closely with the SPG and regularly reports to the Board of Directors.

[5] Enrico Giovannini

4.2 The Role of ICT

As developed in the previous sections, ICT plays a role in many domains related to analytical/statistical information processing. The following list provides an overview of different development domains.

4.3 Development Domains

- Data collection and validation;
- corporate data environment;
- metadata management;
- short-term statistics;
- sectorial statistics;
- multidimensional data analysis;
- time-series management;
- software for economic analysis;
- graphics production;
- desktop tools and front-ends;
- statistical documents & publications (database publishing);
- online OECD statistics;
- electronic data dissemination /data products for sales;
- application-specific requirements;
- A/S computing infrastructure (Hardware and Operating Systems).

An essential responsibility of the OECD Information Technology and Network Service, which cuts across the complete range of business domains, is to provide elaborate a coherent overall architecture in line with the ICT strategy.
Major bricks in the software infrastructure used across OECD analytical/statistical areas today are[6]:

- MS SQL-Server 7 for data storage and cataloguing;
- ORACLE Express for multi-dimensional data-manipulation and 4GL programming[7];
- FAME for time-series data manipulation, graphics, and 4GL programming;
- MS-Excel/Access for common data manipulation;
- Visual Basic and C for programming.

The bulk of processing requirements relates to programmed data manipulation. Users have a need to automate recurrent activities, such as the reception of statistics from Member countries and the preparation of statistical publications. This requirement - bulk data processing - is reflected by the availability of ORACLE Express and FAME, both of which feature a sophisticated, specialised fourth generation programming language.

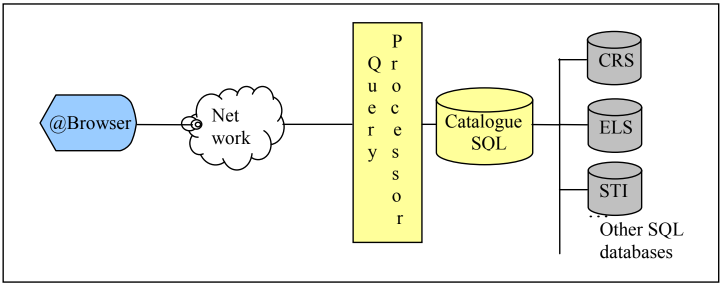
4.4 Initiatives to Date to Develop Statistical Services Based in Internet Technology

In light of the above needs and strategic directions, the OECD has been working on innovative ways to offer new services for the management and dissemination of statistical data and metadata. First steps were taken in 1996 to offer access to specific statistical data sets, in particular the OECD Creditor Reporting System (CRS), which is

[6] In addition SAS is used extensively in one Directorate, particularly for the handling of survey (cross-sectional) data.
[7] ORACLE Express is also used for the storage and management of OECD's International Trade database.

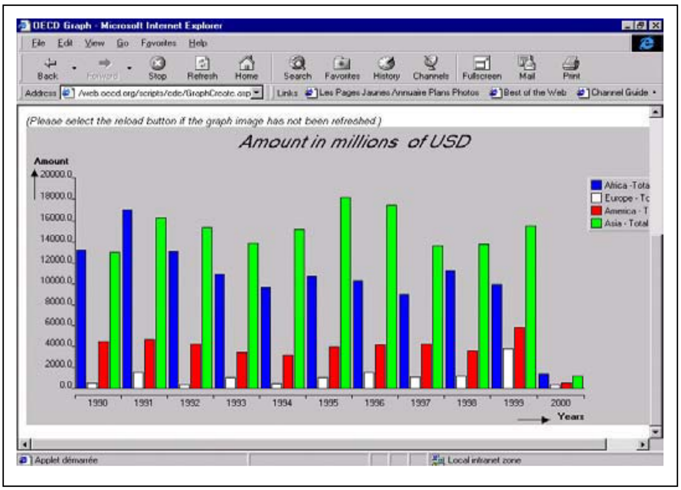
characterised by a relatively high update frequency. A system based on standard relational database technology was devised to offer Intra- and Internet access to dynamically updated data from standard Web browsers. From a technical point of view, this work then evolved into the creation of a "Corporate Data Environment" (CDE) which allows accessing statistical indicators held in different locations from a single entry point using a data catalogue. The visualisation and export of selections, however, remains limited to a single source at a time. Figure 1 illustrates the architecture of this data-warehouse. All databases included in this initiative are SQL Server databases.

Figure 1:. *CDE Architecture Overview*



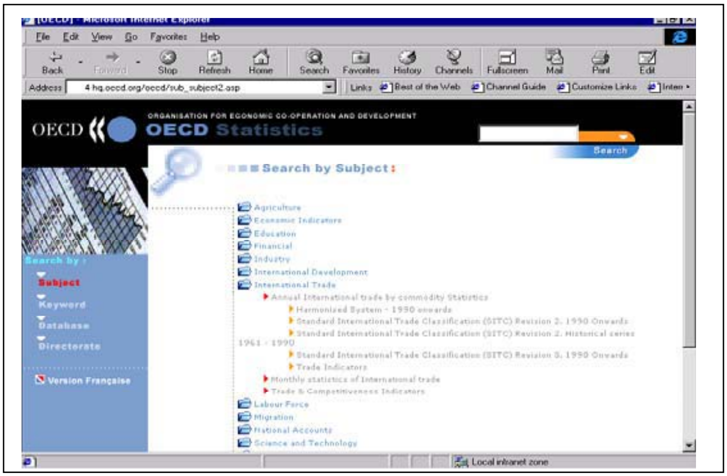
Access to metadata is ensured at both the catalogue and the individual database or data set level. Through its query processor, the CDE features mechanisms to easily find selected statistical data and metadata, carry out comparisons, and compile outputs in various formats. Figure 2 shows a sample output graph from the CDE.

Figure 2.: *CDE sample output graph*



More recently, efforts have concentrated on improving the on-line access to a broader range of official OECD statistics independently of their source database environment. The interactive access to these sometimes rather large data sets has complemented traditional publishing channels/media - see Figure 3. An important consideration in this context was to ensure that the respective data sets are compiled only once for use on different media. This not only allows a reduction of the amount of work and time to market, but also diminishes the risk of inconsistency. Altogether OECD Statistics online cover 14 themes (or subject areas) from Agriculture to Taxation.
Following the standardisation of OECD Electronic Data Products (EDPs) on a single dissemination software – Beyond 20/20 – the OECD started to work with the editor of this software[8] on the functional design for an access mechanism to Beyond 20/20 data files using Internet technology. As with the CDE, all data is stored in a common format. In the first phase of this collaboration, efforts concentrated on the definition of functional and ergonomic aspects. A further standardisation of nomenclatures used in the different data sources was required before releasing a consolidated set of official OECD data on the Beyond 20/20 Web Data Server (WDS). In the context of this work, the multidisciplinary character and the decentralised structure of the OECD very well illustrates the importance for adequate systems support of metadata requirements.

Figure 3: *OECD Statistics On-line today*



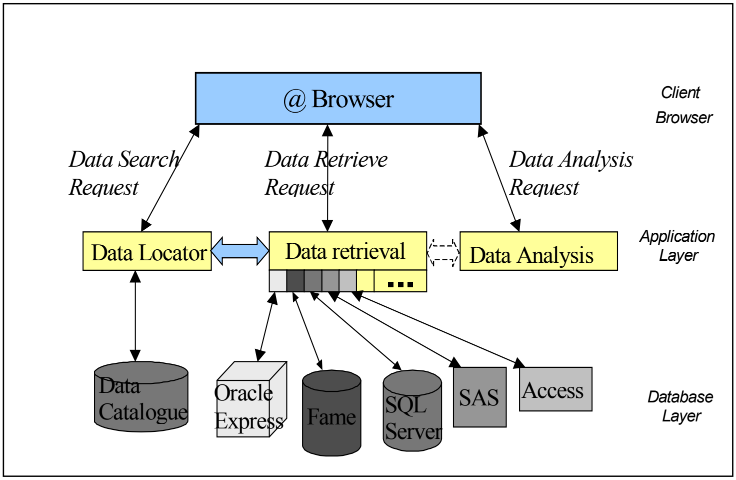
4.5 The Common Browser Concept

Based on the results of the requirements study outlined in section 3, and in line with the OECD Chief Statistician's "New Strategy for the OECD Statistical System", technical solutions for implementing a Common Browser are being studied. Looking at the overall concept illustrated in Figure 4 it becomes clear that there are many commonalities with the CDE architecture. The main difference is the fact that with today's technology it should be possible to build a "broker function" between different

[8] Beyond 20/20 Inc. - © Beyond20/20, Web Data Server (WDS)

types of database-engines. It should be able to provide access to different database sources (types of technology) and to combine the results into composed tables. An important technical consideration is that relational database management system vendors have begun to integrate OLAP (online analytical processing) functionality into their engines. This OLAP functionality is particularly useful for the management of multi-dimensional data objects. This tendency is complemented by the steep increase in speed and of power of modern processors as well IO systems and should allow a tighter integration of different data sources into a common repository. Before deciding on a final architecture and choosing development tools, prototypes are being developed to validate the concept.

Figure 4:. *Common Browser Architecture Overview*



4.6 Design and Implementation Considerations

The following design and implementation considerations for a Common Browser have been identified:

- access to all Directorate statistical databases from a single browser interface available on the Intranet;
- ability to search across databases to locate data on a theme;
- database-level metadata (objectives, person responsible, scope, sources, last update, etc.);
- development of a central data catalogue and dictionary;
- ability to select and combine time series from different data databases ("Shopping Basket" feature);
- easy access to frequently-accessed time series, such as GDP, CPI, Population statistics; the group referred to these series as "OECD Reference Series";
- ability to save selected data in a format suitable for export to analysis software of the user's choice (Excel, E-Views, FAME,..);

- ability to apply FAME's time series analysis and charting functionality to any time series available from the browser;
- databases in the common browser could be accessed in the context of existing structures in order to avoid data duplication and more work for Directorate database managers;
- adoption of a scalable approach: databases could be added on a database by database basis;
- leverage existing browser developments (CDE, STD Browsers, OecdFAME Wizard, WDS…);
- software licensing issues.

4.7 Expected Benefits

- Access all OECD databases using a common interface: "Common Browser" feature;
- minimise the need to install and learn several applications;
- rapid location of data for a single theme from many sources;
- increase the visibility for the most frequently used statistics across the OECD: Reference series feature;
- opportunity and incentive to improve the quality of OECD statistical data, especially metadata;
- minimum set of analysis and charting functionality available from a common browser;
- save selected series in a format suitable for use by other software (FAME, Excel, E-views etc.);
- select and compile sets of series from different databases: "Shopping basket" feature;
- integrate data from different locations integrated into the same presentation: "Virtual database" feature;
- creation of a uniform platform enabling the monitoring of data usage;
- identification of areas of data duplication.

In summary, the objective of the "Common Browser initiative" is for end-users to view data from the relational and specialised database sources together as a coherent database management environment with the help of a central and unique data and metadata catalogue. A middleware layer will be required to dynamically support data search, data retrieve and data analysis requests across this hybrid database architecture based on input parameters and on Catalogue information - see Figure 4. The user will be guided through the panels, based on accesses to the data catalogue and dictionary/glossary. The interface should also offer a set of pre-defined queries for frequently accessed information ("reference series"), as well as the possibility to store queries and to easily export results to other packages. In addition to the interactive mode of access, the middleware should also interface with standard desktop packages, or in programmed mode (through an API) with data manipulation and programming languages (e.g., FAME, Express).

5. Conclusions

The seamless access to statistical data and metadata in support of international policy analysis and decision making is of strategic importance to the OECD. The innovative use of ICT is key to meeting this challenge. The described development strategies are meant to help in the development of statistical information management and services architecture. The objective is to federate "vertical" sets of data and metadata into comprehensive repository through crosscutting catalogue and data search functionality.
The use of modern ICT offerings and the development of adequate implementation concepts are essential to facilitating the management of corporate information and flows, as well as related outreach activities in a multidisciplinary environment. Internet-based information dissemination technologies can not only help meet business priorities in this area, but also allow shortening data collection and publication cycles.
Finally, the implementation of common functions to integrate the collection and initial storage of data and metadata will help to increase timeliness, eliminate errors and duplication, and reduce the reporting burden on Member countries.

Acknowledgements

The preparation of this paper would not have been possible without the valuable input of many colleagues most of which are also longstanding members of the Analytical/Statistical Task Force. Particular thanks go to: Ayse Bertrand, Susan Cartwright, Michel David, Eric Espinasse, Trevor Fletcher, Michael Franklin, Jean-Louis Grolleau, Andreas Lindner, Pascal Marianna, Casper Meyer, Douglas Paterson, Gérard Salou, Cengiz Tarhan, and Colin Webb.

The Data Warehouse: A Modern System for Managing Data

Fred A. Vogel
USDA, NASS, Department of Agriculture,
1400 Independence Avenue. 20250-2000, Washington, D.C., USA
Tel.: 202 6908141, Fax.: 202 6901311, e-mail: fvogel@nass.usda.gov

1. Introduction

There are fundamental forces sweeping across agriculture that are significantly affecting organizations that produce agricultural statistics. Statistical organizations like the National Agricultural Statistics Service (NASS) are increasingly being asked to supply information to aid in making public policy decisions on everything from the transitioning from a planned to a marketing economy to monitoring agriculture's affect on the environment and food safety.

The explosion of the Internet has opened the floodgates of people wanting access to information. The data users are quickly becoming increasingly sophisticated; wanting more data in greater detail, quicker, and with improved accuracy and reliability. The globalization of trade and markets has only increased the need for data of higher quality than ever provided before.

The challenge to statistical organizations is that the additional demands for data are not always accompanied by appropriate resources to satisfy these needs using conventional methodology. Additionally, the changing structure of agriculture to fewer but larger farms is increasing their reporting burden and resulting in declining response rates.

A long-time dream of researchers and analysts has been to have the ability to link historical sample survey and census data for individual farms across long periods of time. The development of improved statistical estimators and analytical tools would be enhanced by easy access to historical data. Data quality could be better monitored if a respondent's currently reported data could be readily linked to what was previously reported. Respondent burden could be reduced by eliminating redundant questions across survey periods and using previously reported data instead. Respondent burden could also be reduced if their data reported for administrative purposes could be used in lieu of survey responses.

The good news is that technology is coming to the rescue. There have been advancements in software and hardware technology that make easy and efficient access to current and historical data a reality.

The following sections will describe the data situation in NASS that led to the need to develop a Data Warehouse. This will be followed by a brief technical discussion about the hardware and software systems that support the Data Warehouse. Then the paper will include a review of how the Data Warehouse is now being used. The summary will include a discussion about the guiding principles the statistical organization needs to adopt along with the high level of management commitment required to successfully implement a Data Warehouse system.

2. The Data Situation in NASS

There are about 2.1 million farms in the United States (U.S.). The size distribution is extremely skewed with the 180,000 largest farms accounting for three-fourths of the annual sales of agricultural products. Only about 5,000 farms account for one-fourth of the sales.

Over 400 reports are published annually by NASS. These statistics are inclusive in their coverage of crop and livestock production on a weekly, monthly, quarterly, semi-annual, and annual basis. Some reports contain data on economic and environmental indicators. The official estimates in most of these reports result from a system of sample surveys of farm operators. Several hundred surveys a year are conducted that cover over 120 crops, 45 livestock species and associated economic and environmental items. In addition to the ongoing survey program, the results of the Census of Agriculture are published at five-year intervals.

The skewed size distribution of farms results in many of them being included in multiple surveys during a year. The situation is compounded by the very nature of the NASS statistics program. For example, one purpose of the crops estimating program is to provide data for forecasting purposes. Before the crop planting period, a sample of farmers is asked to report the area by crop they intend to plant in the coming season. This is the basis for the *March Acreage Intentions* report published in late March each year. This is followed by another survey in June after crop planting to obtain crop areas planted and expected areas to be harvested for grain. Monthly surveys are then conducted during the growing season where sampled producers are asked to report the yield they expect to produce. After harvest, another sample of producers provides data on areas actually harvested and quantities produced. Four times a year, sampled producers are asked to report the quantities of grain stored on the farm. Table 1 provides examples of questions asked by survey period.

Table 1: *Typical Questions asked of Producers*

Typical Questions asked on a Sample Survey	March	June	August	September	September-Through November	December
How many acres of ____ expect to plant?	X					
How many acres of ____ have been planted?		X				
How many acres of ____ expect to harvest?		X	X			
What do you expect the yield of ____ to be?			X		X	
How many acres of ____ did you harvest?						X
How many bushels of ____ did you harvest?						X
How many bushels of ____ are in storage?	X	X		X		X
How many pigs born the last 3 months?	X	X		X		X
How many sows are expected to farrow the next 3 months?	X	X		X		X

The sample surveys are designed to provide both direct estimates and ratio estimates of change. To reduce respondent burden, the samples are also designed so that producers rotate into and out of surveys based on replicated-rotating sample designs. Some farms, because of their sheer size are included with certainty in every sample.

A brief summary of the issues and problems that provided the impetus to develop a Data Warehouse follows.

- Processing systems for various surveys have generally been determined by the size of the application. As a result, processing platforms include a mainframe, Unix platform, and personal computers connected into Local Area Networks. Software includes SAS, Lotus, X-Base, and other systems. Therefore, it has been very difficult or nearly impossible to link data across survey systems, because of the variety of software and hardware platforms being used. Even when the data were in one software, such as SAS, separate files were maintained for each state and survey; resulting in thousands of separate files being created each year.

- Many farms, especially large farms, are selected in recurring sample surveys. As a result, they can often be asked questions about items they previously reported. The NASS surveys are voluntary, thus producers do not have to report every time, and some do not. Imputation for missing data would be greatly enhanced if previously reported information could be readily used.

- Data analysts find it difficult to access the myriad of systems for ad-hoc analysis without the aid of computer programmers who are often too busy or reluctant to provide assistance. As a result, many questions raised during the analysis of survey data go unanswered. It is also difficult to link historical and current survey and census data to improve estimation and imputation procedures.

For these reasons and others, NASS made a strategic decision in late 1994 to implement a data base that would contain all individual farm level reported data across time.

3. What is a Data Warehouse?

The Data Warehouse, using the NASS situation, is a data base system that contains all "historical" individual farm level data from the *1997 Census of Agriculture* and major surveys from 1997 to present. Also, the Data Warehouse now contains data from "current" surveys so that one-stop shopping of current and previous data is provided. When the 2002 Census is being conducted, farm level reported data will be added to the Data Warehouse during the day, every day, to facilitate data review and analysis.

A guiding principle in the choice of the data base and in its development was that the data be separated from the application systems. A driving force was to make the data easily accessible by all statisticians using simple access and analysis tools without requiring intervention of programmers or systems analysts. Another system requirement was that it be easy to track data across surveys/censuses and across time – a necessity to optimize use of historical data.

The data base design needed to support fast data loading and speedy data access in contrast to transactional processing where the emphasis is on adding, deleting, or updating records rather than accessing them. The Data Warehouse must provide rapid read-only access to large chunks of data records across surveys and across time.

The Data Warehouse "engine" at NASS is the Red Brick Decision Server, provided by the Informix Company. It is operated on an IBM Unix Box with four CPU's, expandable to eight. There are 500 gigabytes of disc space and over 900 million rows of data currently in the data table.

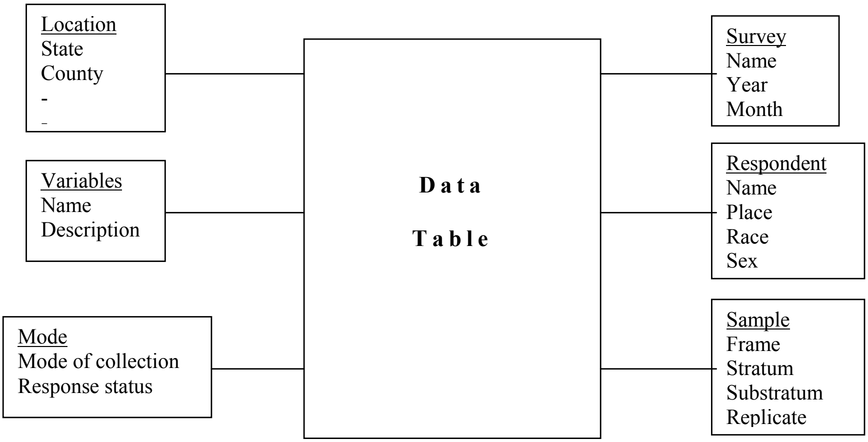
A team of eight people (two full-time and six as advisors) developed the data models and implemented the system. Work on the project started in July 1996, and equipment and software were purchased in December 1997. The data collection for the *1997 Census of Agriculture* started on January 1, 1998, and by April 1998 the Data Warehouse was populated with incoming census data. The Data Warehouse played a central role in the analysis of the census returns compared with previously reported survey data from 1997. There are currently five people working on the Data Warehouse, which includes one database administrator.

4. How does it work?

The intent of the Data Warehouse was to provide statisticians with direct access to current and historical data and with the capability to build their own queries or applications. Traditional transactional data bases are designed using complex data models that can be difficult for anyone but power users to understand, thus requiring programmer assistance and discouraging ad-hoc analysis. The database design results in many database tables (often over 100 tables), which result in many table forms when querying, so query speed is often slow. Again, technological developments in the industry came to the rescue. The goal to provide end users simple access and analysis capabilities quickly led to the development of the dimensional modeling concept. Dimensional modeling is the primary technique for data bases designed for ad-hoc query and analysis processing.

In effect, the data view involves two groups of tables including data and the metadata. The metadata tables contain the information that identifies, explains, and links to the data in the data table. Table 2 shows the schematic of the overall dimensional model designed for the NASS Data Warehouse. Notice that there are only seven tables in the Data Warehouse – one data table and six metadata tables.

Table 2: *The Data Warehouse Dimensional Model*



The central data table contains the individual data responses for each question for every survey or census from 1997 forward, when available, a weight is also provided that can be a non-response adjustment or sampling weight so that direct estimates can be readily derived. The six surrounding tables contain the metadata describing the survey or census response. These six tables were designed in a way that reflects how the end user or commodity analyst views the data, which is by a given variable (corn acres planted) by location (state) by survey (June acreage) by sample frame (stratum) by respondent, and by mode (method of collection or response). Users simply browse the six metadata tables to define what data they want to access from the data table for their analysis.

Every statistician in NASS can access the warehouse in a read only format. NASS has implemented an easy-to-use query tool (BrioQuery) purchased from Brio Technology. Using this query and analysis tool and the dimension table logic, statisticians build their own queries to generate tables, charts, listings, etc. Canned applications are also prepared for analysis situations that occur frequently.

Some simple examples of ad-hoc queries developed by users follow:

- Provide all historical data as reported over time by one hog producer. The time period included every quarterly survey for the last three years.

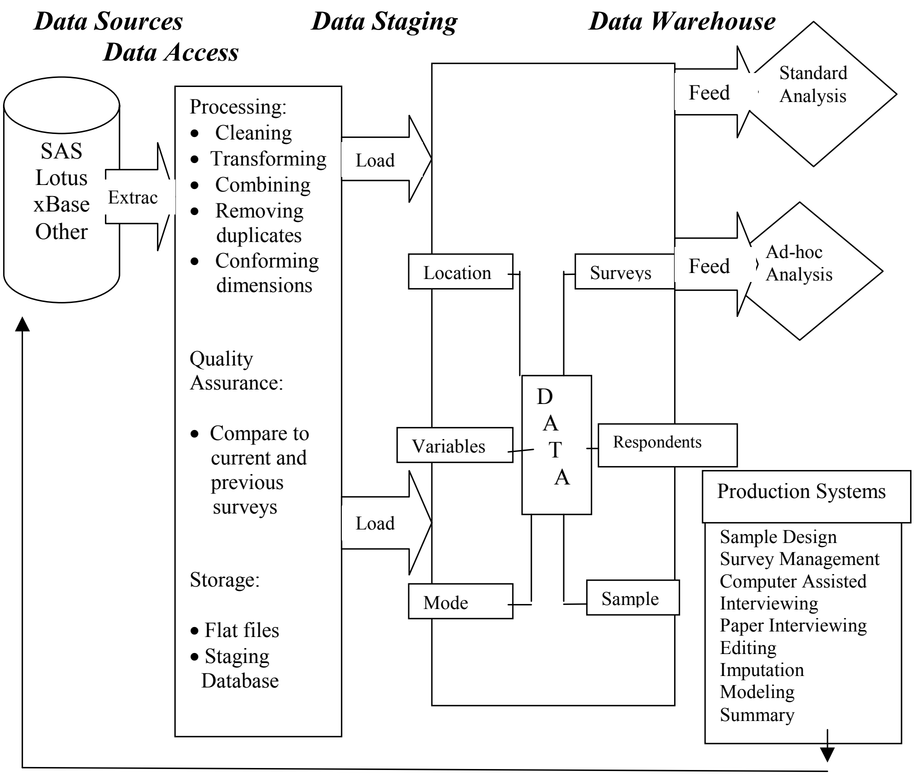
- Provide all data over time for the 100 largest hog producers in the U.S. arranged from the largest to the smallest.

- Provide historical data for selected variables for all non respondents to the current survey.
- Obtain a preliminary sample weighted total of hog inventory for the current survey.

- Obtain number of farms, acres, and production for all farms in the U.S. reporting corn yields of 10 bushels per acre or less.

These applications can be processed and returned to the end user in seconds. The ability to slice and dice the data is limited only by the statisticians' desires and ingenuity.
Table 3 shows the overall data flow. Remember that one of the original guiding principles was that the data be separated from the applications. Table 3 illustrates how the survey systems load data into the warehouse and how the warehouse feeds data into the analysis and production systems.

Table 3: *NASS's Architecture*



5. What we learned

Senior management support is crucial, because the development of a Data Warehouse containing all census and survey responses with metadata in one place **AND** easily accessible to everyone strikes a blow to a traditional organizational structure and ways of doing business. There is great resistance to change by many, so senior management must realize its strategic value and help sell its merits to others in the organization.
The Information Technology (IT) personnel may actively oppose such ideas and view them as a threat or not understand the analysts need for data. Some reactions NASS heard from its IT managers were:

> "users do not know what they want"
> "users will not be able to do their own queries"
> "it will not be compatible to "our" system"
> "can't pull that data together now"
> "why do you need to have previous data readily accessible"
> "no one will use it because they're too busy"

It needs to be said that the primary reason NASS has a Data Warehouse is because it was developed by statisticians in the program divisions who understood the value of data. NASS's Data Warehouse was developed and implemented without IT support. Even now, responsibility for the Data Warehouse resides in a program division, not the IT Division. However, to promote more support in the future from the IT Division, one of the two designers of our Data Warehouse has been transferred to lead our IT Division.
The end users in the past (unless they transformed themselves into power users) were dependent on the IT professional to get at their data. The IT professionals often determined what data platforms and software would be used for an application with little thought being given to the end user. The philosophy was "Tell me what you want, and I will get it for you."
End users need to be involved at the design stage so that the dimensional models reflect their way of viewing the data. Too rarely, end users are brought in at the end of the development chain, then held hostage by the system provided to them. For the Data Warehouse effort to succeed, senior managers need to ensure that end users are fully engaged in the developmental process.
The desire for simplicity should prevail. There are simple solutions to complex problems, and the end user view often points the way to the solution.

6. What is in Store for the Future?

The 400+ reports NASS publishes each year are not close to capturing all of the information embedded in the census and survey data files. As the statisticians become more capable in their use of the Data Warehouse, they will be able to "mine" the data for new relationships or conclusions that will enhance the reports.
Another major accomplishment would be to store all of the summary level information in the 400+ reports across time in the Data Warehouse. Then data users could create

their own data tables rather than being limited to the basic formats now being provided in the official releases.

A huge achievement will be the capability to use historical information to "tailor" questionnaires for each respondent. Respondents will not be asked to provide information reported earlier, but only to update the information as needed. A recent development was to design a process to load current survey data into the Data Warehouse as soon as it passes through a preliminary edit. Thus, current data, and estimation can benefit from the connection to historical data at the record level.

7. Conclusion

The development of the Data Warehouse concept ranks as one of the most significant technological and methodological NASS has ever achieved. The entire sequence of events required to produce official statistics will be affected by the warehouse concept and technology. This new technology is also rocking the very foundations of the organization's way of doing business. The hardware and software systems can be developed independently of the platform. This simplifies the developmental process. The most significant change is, that the data management and data use activities revolve around the end user.

Acknowledgements

The author was the executive sponsor of the Data System 2000 Strategic Initiative Team that initiated the development of the Data Warehouse at the National Agricultural Statistics Service (NASS). Input for this paper came from presentation material prepared by Jack Nealon, Mickey Yost, and Jim Cotter from NASS. Without their dedication, passion, persistence, and zeal, the NASS Data Warehouse would not be a reality today.

First Agricultural Census in China: Data Access and Dissemination

Antonio Giusti
Dipartimento di Statistica, Università di Firenze
Viale Morgagni, 59 – I 50134 Firenze, Italy
e-mail: giusti@ds.unifi.it

Abstract: The First National Agricultural Census (FNAC) of the People's Republic of China represents the most demanding statistical computing task ever accomplished in the world. SDA software, a set of computer programs for the documentation and Web-based analysis of survey data, was used to permit the use of census data directly by users and to supply the researchers with an instrument for a complete, safe and adequate access to the data to perform statistical analysis. The use of SDA is very simple and data processing time is not related with the computational power of the computer used to navigate.

Keyword: Agricultural Census, China, Dissemination, Internet, SDA.

1. Introduction

The First National Agricultural Census (FNAC) of the People's Republic of China (PRC) represents the most demanding statistical computing task ever accomplished in the world. To understand this statement just consider several aspects of the FNAC: more than 7,000,000 enumerators used, about 220,000,000 forms filled, 220 Gigabytes of data entered (about 80 Gigabytes of compressed files), and so on. No census in the world (including the other censuses in China - population and economic) comes near these figures.
A first characteristic of the FNAC is the heavy implementation of new technologies, thanks to international assistance (above all the support of four FAO Projects: GCP/CPR/006-010-020-025/ITA), and also to the considerable development of PRC.
Technical and organizational aspects were particularly cared for: intensive use of computer systems, with more than ten years spent on the training of National Statistical Officers (both on hardware and software[1]), and the construction of a well-organized geographical network for the enumeration activities and data processing.[2] During the pilot censuses (in 1991, 1995, 1996) different software products were tried for data entry, checking, and imputation: IMPS (developed by Bureau of Census, USA), Blaise (developed by CBS Statistics Netherlands - a special version partially translated in

[1] Training courses on hardware regarded personal computers, servers, Unix machines, networks, maintenance, while training courses on software dealt with operative systems (MS/Dos, Unix, Windows, Windows NT), programming languages, spreadsheets, word processors, databases, and statistical software (especially SAS).
[2] Many data processing centres were instituted in different Provinces, Municipalities and Autonomous Regions. Each centre was provided with servers, personal computers, printers, network facilities, etc.. These centres played a very important role during the preparation and the conduction of the FNAC.

Chinese was used). Other packages were evaluated but not implemented. In total, between the beginning of the training program to the data analysis stage, about 5,000 computers were used. A relevant characteristic of the FNAC was the introduction of the Optical Character Recognition (OCR) to conduct data entry for the "Rural Household Holding" forms (about 210,000,000 in PRC). 573 scanners were used for census data acquisition.[3]

Checking and imputation activities were performed at different levels, with much use of automated checking. A nation-wide quality check on OCR data entry was also conducted, using a sampling scheme.

The data processing was organized at three levels: at national, provincial and prefecture level; at each level raw data were used.[4]

In this way census data are available for analysis at different levels in any Prefecture, Province and National Bureau of Statistics (NBS). The NBS Computer Centre developed a database using Oracle. The database includes three kinds of data: micro data, meta data, and macro data. Micro data include all data collected during the census; meta data will comprise classifications and dictionaries; macro data will include summary statistics and tables. The client/server architecture was chosen. This database was reserved for internal use only. For a more detailed description of this subject see Giusti and Li, (2000).

A database with a 1% sample data was also completed.[5]

In section 2 of this paper we will describe data access and dissemination of the FNAC, while some general characteristics of SDA are presented in the following section. In section 4 we will show a sample session using SDA. A final section is devoted to the conclusions.

2. Data access and dissemination

Data dissemination of the FNAC was carried out using both manual tabulation and computer data processing. Manual tabulation was used to quickly release information on general results: the National Statistical Bureau published five communiqués at the national level. These communiqués are also available in the Web site of the NSB. At the same time, all provinces released their own communiqués.

Computer processing data dissemination concerns macro data (tables and graphics), micro data (files) and interactive queries (to produce tables and files). Data dissemination was completed by using traditional approaches (basically, paper publications). Advanced methods, such as Web site, CD-ROM and on-line databases, are under implementation.

[3] Due to the OCR introduction, available data entry software was not used. The PRC National Bureau of Statistics Computer Centre developed a software program for data entry, checking and imputation, including: address code management, optical data transfer, data entry, editing (verification), and tabulation. This software was realised both for DOS and UNIX operating systems. This software was used both for OCR and keyboard data entry.

[4] To give an idea of the administrative entities involved, this was the specification of the administrative organisations in PRC in 1996: 22 Provinces, 4 Municipalities, 5 Autonomous Regions, 1 Special Administrative Region, 335 Prefectures, 2,849 Counties, 45,484 Towns or Townships, 740,128 Administrative Villages.

[5] A systematic sample was selected from the list frame of households, sorted by province, prefecture, county, town, village and household number.

Using data processing results, several books with census data were published in 1998 and in 1999, such as "Abstract of the First National Agricultural Census in China", "Facts from the First National Census of Agriculture in China", in English and Chinese. At the same time, all of provinces also published the census results in Chinese.

The evolution of data processing and the increasing diffusion of Internet had a very important impact on data dissemination activities carried out by National Statistical Institutes (NSIs). At present, many NSIs disseminate data in electronic form, both using optical devices and network facilities.

Utilising the new technologies, data release can be done in several ways, depending on the type of data to be disseminated (micro or macro data).[6] Micro data files can be released on CD-ROM or allowing the file download with FTP (File Transfer Protocol) or other protocols, usually from the NSI Web Site. Macro data (tables, graphical representations, reports, etc.) can be disseminated in the same ways but paper publication is still the preferred modality.

A more advanced method, to release both micro and macro data, is through an interactive system that allows the user to ask for specific analyses. Some NSIs allow user to produce tables and to download micro data files. We can define this new tool "interactive Web dissemination".

With the cooperation of NBS Computer Centre, the census results have been disseminated by using these advanced methods of data dissemination. For example, the "Abstract" was released in CD-ROMs, some other publications, tables and micro data files were published on the Web site of the NBS[7].

In order to permit the use of census data directly by users and supply the researchers with an instrument for a complete, safe and adequate access to the data to perform statistical analysis, the NBS decided to release the 1% sample data by using the World Wide Web.[8]

To realize this task, we considered experiences made by some NSIs (Bureau of Census, Statistics Canada, ISTAT, etc.) and the new technologies available for data management (data warehouse, multidimensional data management, and so on).

After an accurate evaluation, it was decided to use SDA (Survey Documentation and Analysis): a set of programs developed and maintained by the Computer-assisted Survey Methods Program (CSM) of the University of California, Berkeley. SDA characteristics will be presented in the next section.

For this purpose two National Technical Officers had two weeks period of training in Florence. During this time they prepared a preliminary experimental data set using some records from the FNAC. The work was accomplished using Unix version of SDA.

In July all 1% sample files from A601[9] were implemented at FASC (the Food and Agricultural Statistical Centre of NBS) using SDA in Windows NT environment (Giusti, 2000).

In installing SDA we considered disclosure issues. Some experiments were carried out on the 1% sample database, using some general principles accepted at international level.[10] The 1% sample data is already released on the web.

---

[6] In electronic data release very often micro and macro data files include metadata.

[7] The address of the English version of the NBS Web site is: http://www.stats.gov.cn/english/index.html

[8] For more details on the sample see Pratesi (2000).

[9] The Rural Household Holding form, divided in six sections.

[10] The first experiment was made at Provincial level obtaining as a result the possibility to have one identification every 59 million inhabitants. The second experiment, at county level, showed a possibility

3. SDA general characteristics[11]

SDA software is a set of computer programs for the documentation and Web-based analysis of survey data. There are also procedures for creating and downloading customized subsets of datasets. The software is developed and maintained by the Computer-assisted Survey Methods (CSM).[12]
Data analysis programs are designed to be run from a Web browser. SDA provides analysis results very quickly -- within seconds -- even on large datasets; this is due to the method of storing the data and to the design of the programs.
SDA features:

1. Browse the documentation for a dataset or a questionnaire:
   - Introduction files, appendices, indexes to variables;
   - Full description of each variable.
2. Data analysis capabilities:
   - Frequencies and crosstabulations;
   - Comparisons of means (with complex std errors);
   - Correlation matrix;
   - Comparisons of correlations;
   - Regression (ordinary least squares);
   - List values of individual cases.
3. Create new variables:
   - Recode one or more variables;
   - Treatment of missing data;
   - Compute new variables;
   - List newly created variables.
4. File handling:
   - Make a user-specified subset of variables and/or cases, and download data files and documentation;
   - ASCII data file for the subset;
   - SAS, SPSS, or Stata data definitions;
   - Full documentation for the subset.
5. Other Features under Development:
   - Logit and probit regression;
   - Interface in various languages.

Online help for SDA 1.2 analysis programs is available on the Web. For each program, an explanation of each option can be obtained by selecting the corresponding word highlighted on the form.
In particular, SDA Frequencies and Crosstabulations Program generates the univariate distribution of one variable or the crosstabulation of two variables. If a control variable is specified, a separate table will be produced for each category of the control variable.

---

to have one identification over 3 million inhabitants. These values are very low respect to many NSI rules. So we can conclude that, disabling download option of SDA, disclosure problems can be avoided. For more details see Franconi (1999).

[11] More information on this topic can be found in the SDA Manual (version 1.2), on line at the site: http://sda.berkeley.edu.

[12] CSM also develops the CASES software package, for telephone and self-interviewing.

After specifying the names of variables, it is possible to select the display options you wish: these affect percentaging (column, row or total), text to display, and statistics to show. Using filter variable(s), some cases are included in the analysis; others are excluded. Cases are given different relative weights, using weight variable. Specifying both a row and a column variable, a package of bivariate statistics is generated (Pearson's Chi-square, and the Likelihood-ratio Chi-square, each with its P-value, Pearson correlation coefficient and Eta, Gamma and Tau). Specifying a row variable only, a package of univariate statistics is generated (the mean, median, mode, standard deviation, variance, the coefficient of variation, the standard error of the mean and its coefficient of variation).

With version 1.2, using many programs (including Frequencies and Crosstabulation), the table cells can be colour coded, in order to aid in detecting patterns. Cells with more cases than expected (based on the marginal percentages) become redder, the more they exceed the expected value. Cells with fewer cases than expected become bluer, the smaller they are, compared to the expected value. The transition from a lighter shade of red or blue to a darker shade depends on the magnitude of the T-statistic.

The sample design can be specified for each study when the SDA dataset is defined; this definition allows SDA to compute the standard errors of means for complex samples.

With SDA it is also very easy to create new variables using basic or composite expressions (with If / Else if / Else). In expressions, arithmetic and logical operators, arithmetic, random distribution, statistical and trigonometric functions can be used.

4. A sample session using SDA

The use of SDA is very simple and data processing time is not related with the computational power of the computer used to browse the web site.

To use the 1% sampling data of the PRC FNAC we can start with the page: www.stats.gov.cn/english/index.html. In the page is present the reference to the interactive electronic data release used for the FNAC. From this page, after another mask that allows the access to the project documentation, it is possible to directly reach the study and action selection page (see figure 1).

The study selection is necessary to access one of the six sections of the form (to access the file that we would like to analyse). The action selection permits to choose among: browsing codebook, making frequencies or crosstabulation, performing comparison of means, calculating correlation matrix, using multiple regression program, and listing values of individual cases.

We suppose to know the file structure (contents and characteristics of each variables recorded in the six sections) that is necessary to go on with the requests. Otherwise, using the action “Browse codebook” it is possible to get the name, the label, the categories and the “question text” of each variable.

Supposing to choose “Frequencies and crosstabulations”, we go to the mask in figure 2. In this example the data section used is the “General Information of Household”.

In figure 2, we can see the field filled to ask a crosstabulation using as row variable the “province” (or equivalent administrative level) and as column variable the “number of persons in the household”. We have not defined any weight, recode or class definition. In few seconds (normally 3 or 4) it is possible to receive the results (see figure 3), despite the relevant number of processed records (2,138,275). After the description of the variables used, the table is presented. The use of colour-coding allows a quickly identification of the most common patterns.

Figure 1: *Study and action choose*

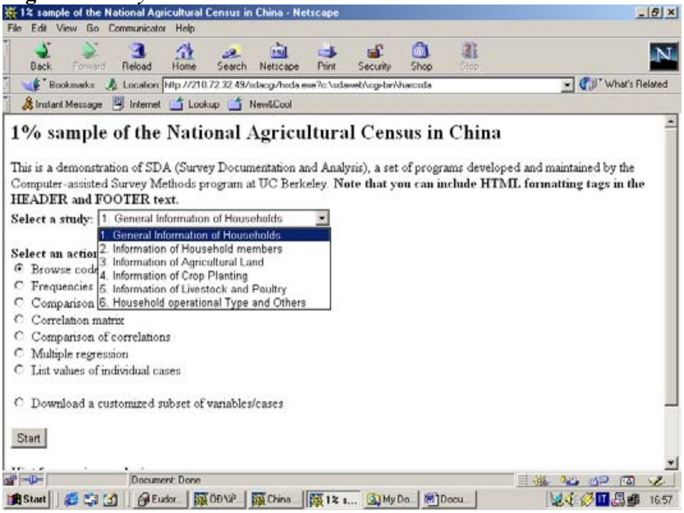


Figure 2: *Tables program*

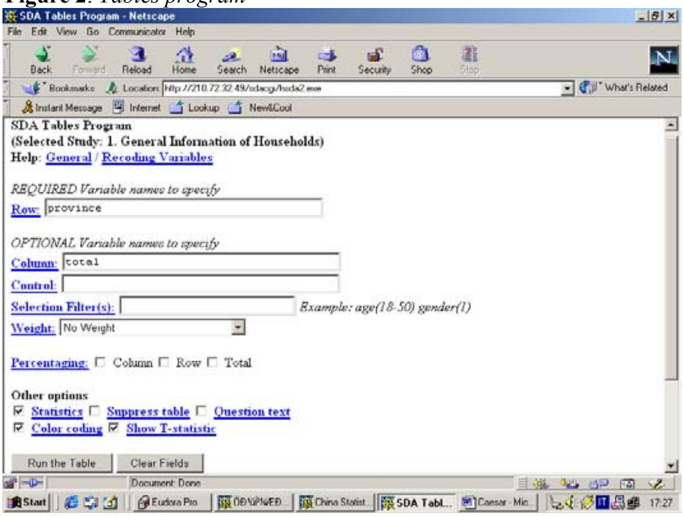
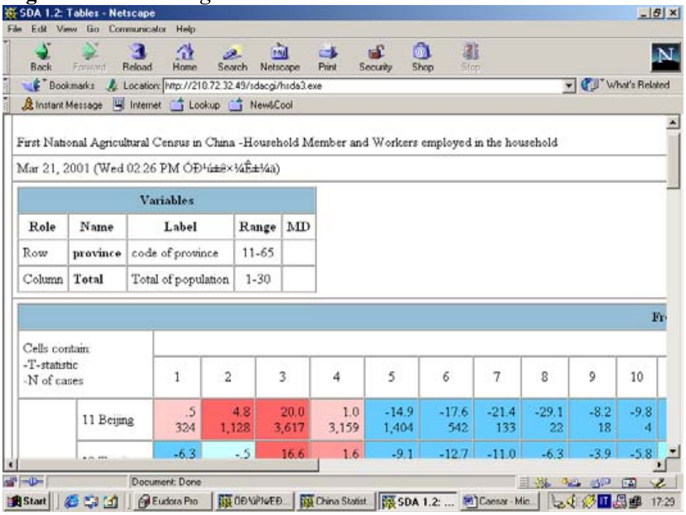
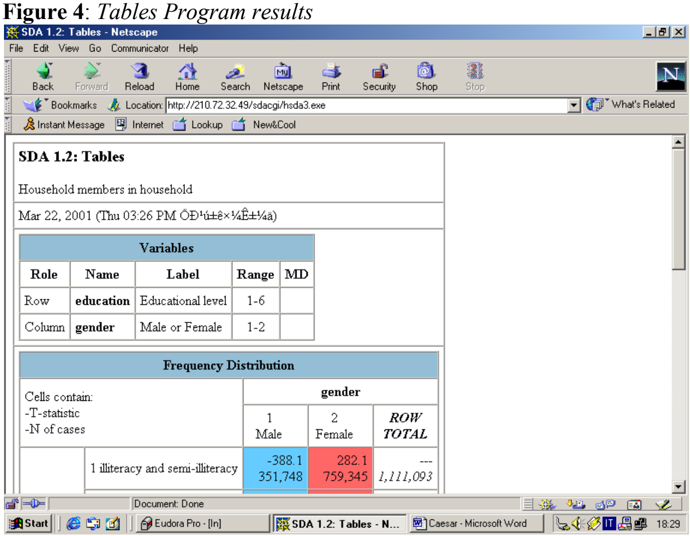


Figure 3: *Tables Program results*



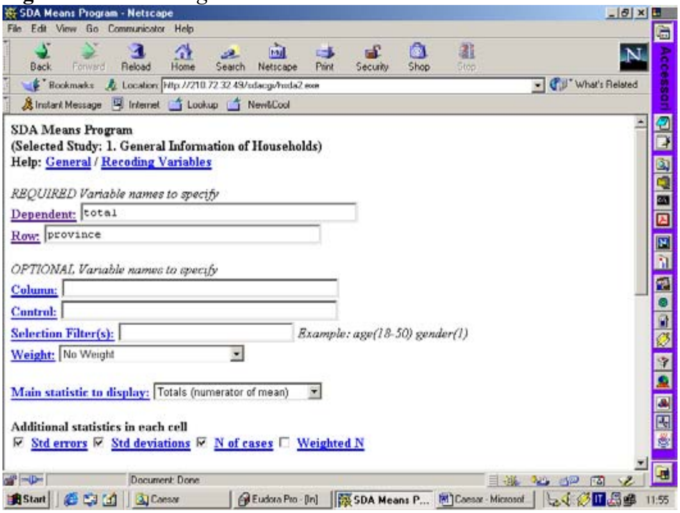
Using the option in figure 1, we can change the study, to analyse "Household members data". After, using the mask in figure 2, we ask a crosstabulation using as row variable the "education" and as column variable the "gender". In this situation 7,826,390 records will be used to obtain the results. Data processing time remains well below the 10 seconds (figure 4).

Figure 4: *Tables Program results*



The last example is devoted to the calculation of the "total population" by "province". In figure 5, we present the request, using the Means Program, while in figure 6 we show the results (top and bottom parts).

Figure 5: *Means Program*



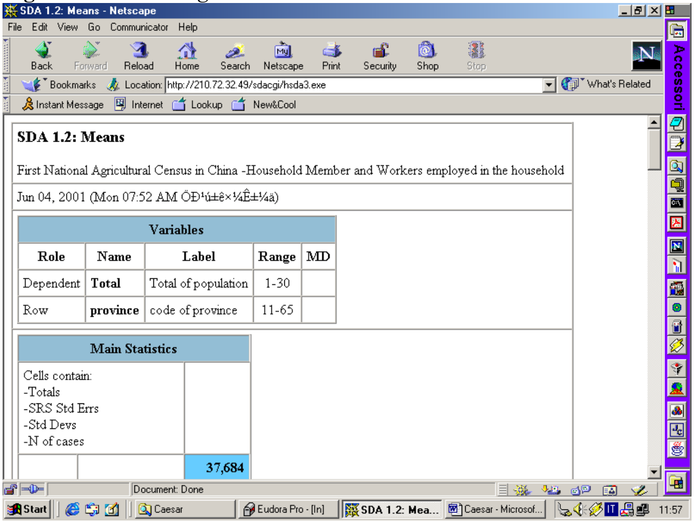
5. Conclusions

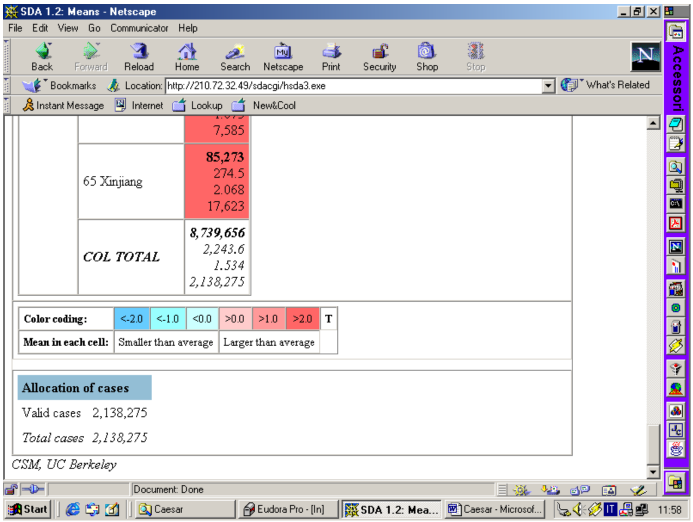
Data processing and data dissemination represent an important part of the PRC FNAC activities. In order to permit the use of census data directly by the users and to supply the researchers with an instrument for a complete, safe and adequate access to the data to perform statistical analysis, the NBS decided to release a 1% sample data by using SDA on the World Wide Web. With SDA it is possible to produce codebooks and to analyse the census data from any remote personal computer. The SDA experiment demonstrates the interest of China NBS, FASC and FAO in new technology implementation in dissemination strategy through Internet.

Now, the main idea is to integrate in the FNAC dissemination as many as possible new technology tools to increase the accessibility and the utilisation of a so precious source of information.

More advanced techniques of data analysis, including some based on interactive GIS (Geographical Information System), should be used widely in the future. Some maps will be released on the Web. Data integration with other sources of information will be planned. The integration of continuous updating of agricultural statistics in the Web NBS site will represent a very important and advanced tool for the knowledge of agricultural information of the PRC.

Figure 6: *Means Program results*





We will conclude this paper with a last consideration, devoted to a problem that apparently could be considered only terminological. In this presentation, the term "data warehouse" and "data warehousing" are not used. This is not due to a lexical preference: in our opinion a data warehouse is something different from this and other similar experiences.

A data warehouse is not just a file or a database used for dissemination purposes: is the result of some transformation of one or more operational files and/or databases in a static database (with temporal information) to be used for statistical information selection (Giusti and Martelli, 2000).

In an enterprise data warehousing is an important option to create databases for decision support. For a NSI the situation is very different because the construction of statistical databases for data dissemination is the main purpose of data processing activity. A census data file is not an operational database, and a survey data file is not subjected to transactions, after the editing and imputation steps.

NSIs deal with an information system in which the production of statistical databases is the main goal; their situation is completely different from other enterprises where the primary scope is not the dissemination or the use of statistical information.

References

Abstract of the First National Agricultural Census in China (1998), China Statistics Press, NACO, Beijing.

Facts from the First National Census of Agriculture in China (1998), China Statistics Press, NACO, Beijing.

Franconi L. (1999), *Report of the visit of the FASC by Luisa Franconi, (July 1999)*, FAO internal report.

Giusti A. (2000), *Report of the visit of the FASC by Antonio Giusti, (July - August 2000)*, FAO internal report.

Giusti A., C. Martelli (2000), "Basi di dati e processo di produzione dei dati statistici", in: L. Fabbris (a cura di), *Rilevazione di dati assistita da computer*, CLEUP. Padova.

Giusti A., Li Wenhai (2000), "Data processing and Dissemination", in: *International Seminar on China Agricultural Census Results*, FAO, Beijing.

Pratesi M. (2000), *Notes on the dissemination of the 1% Sample Data from the Census of the Agriculture in China*, FAO internal report.

Summary Report on the First National Census of Agriculture in China (2000), China Statistics Press, NACO, Beijing.

Theses of the Data Entry Technology by OCR of the China's First National Agricultural Census (1998), Haichao Press, Computer Center, Beijing.

WEB sites:

http://210.72.32.49/default.html
http://csa.berkeley.edu:7502/
http://www.stats.gov.cn/english/index.html

7 June 2001

Invited Paper Session

ADVANCED SURVEY DESIGN AND ESTIMATION ISSUES

Chair: G. Cicchitelli

Optimal Spatial Sampling Strategies for Agricultural and Environmental Data[1]

Giuseppe Arbia
Department of Sciences, Faculty of Economics, University "G. D'annunzio"
Viale Pindaro, 42, 65127 Pescara, Italy
e-mail: Arbia@sci.unich.it

Giuseppe Espa
Institute of Operational Research, Faculty of Economics, University of Trento
Viale Inama, 5, 38100 Trento, Italy
e-mail: Gespa@gelso.unitn.it

Abstract: The aim of this paper is twofold. The first aim is to review spatial sampling designs in the presence of spatial dependence. We discuss the use of traditional sampling strategies when collecting spatial data and we motivate the need for *ad hoc* techniques to achieve higher levels of efficiency. We refer to these techniques as "contextual designs". We review some of the techniques proposed in recent years and we argue in favour of a draw-by-draw strategy that appears to be a convenient way of achieving optimality by reducing the computational burden. The second aim of the paper is that of providing a (tentatively) exhaustive bibliography on the topic.

Keywords: Agricultural surveys, Contextual designs, Environmental assessment, Optimal designs, Sequential sampling, Spatial correlation.

1. Background

The application of statistical sampling theory to problems involving spatially distributed agricultural and environmental data has a long tradition dating back to Hasel (1938) and Mahalanobis (1940). Ten years later Matern (1947) and Finney (1948, 1950, 1953) discussed the relative efficiency of random sampling versus systematic sampling for estimation problems in agriculture and forestry. In these years we can find also a series of mathematical contributions to the problem. Cochran (1946) presents the one dimensional problem and Quenouille (1949) and Ghosh (1949) generalize it to the two dimensions. In the fifties Das (1950) Milne (1959) and Williams (1956) treat in particular random and systematic sample designs and derive their properties in two dimensions. To these years is also dated a mathematical proof that systematic sampling outperforms random sampling in terms of the variance of error of the sample mean. (Zubrzycki, 1958). Dalenious et al. (1961) seek the design that minimizes the estimation variance in plane sampling. A change in the direction of the studies occurred after the formalization of multidimensional random fields due to Yaglom (1957, 1961, 1962)

[1] Partial financial contribution of MURST 60% of Chieti University and of the Italian National Research Council # 99.01511.CT10 is also gratefully acknowledged.

Matern (1960, 1986) Whittle (1952, 1963), and after the introduction of the geostatistical approach by Krige (1951), and Matheron (1962, 1963, 1965). In the seventies and eighties there are numerous contributions to spatial sampling in agriculture, forestry and many other applied fields (like e. g. mining, geography, meterology, hydrologic sciences). A precious source of this literature is contained in Barnes (1989). Holmes (1970) examines sampling design with applications to geography, Payandeh (1970) addresses sample design efficiency in forestry. Viera et al. (1981) discuss applications in the soil sciences, and analyse the relative effects of sample design on the estimation accuracy. A wide literature can also be found in hydrology (see Rodriguez-Iturbe and Mejia, 1974; Bras and Rodriguez-Iturbe, 1976; Bras and Colon, 1978; and the special issue of the Water Resources Research, 1979).

Despite the amount of published work in the literature the idea of including contextual information in the sampling design is only a recent one. In this paper we will review the traditional solution to optimal sampling design to collect agricultural and environmental data and discuss their limits and disadvantages (Section 2). In Sections 3 and 4 we will review a set of techniques proposed in recent years to include contextual information in the design. We will distinguish between probabilistic designs that use the context (to be contrasted against simple random or stratified random) (Section 3.1) and non-probabilistic designs (to be contrasted against systematic sampling) (Section 3.2). Section 4 contains some conclusion and a selection of topics for further research in the field.

2. Traditional solutions to areal sampling design

Traditional textbooks (Hansen et al, 1960; Kish, 1965; Mailing, 1989; Webster and Oliver, 1990) describe various area sampling techniques. We can distinguish between probabilistic designs (like simple random, stratified random, two/three stage sampling etc.) and non probabilistic designs (like systematic). Mixed strategies are also adopted when using for instance, unaligned systematic designs.

When employing these techniques the selection of areal units is made as if the generating process which underlies the observed values, is constituted by a sequence of random variables independently and identically distributed. However a distinctive feature of spatial data is that of being spatially dependent, a problem fully recognized at the stage of data analysis and statistical modelling (Cliff and Ord, 1981; Ripley, 1981). Dependency implies that "data of geographic units are tied together like bunches of grape, not separate like balls in an urn" (Stephan, 1934) so that they cannot be thought of as randomly generated from the classic urn model. Thus the generating process underlying the sample is constituted by a sequence of random variables that are no more i.i.d.

Since in geographical studies "adjacent units are often more alike than units that are far apart" (Cochran, 1963; p.96) it is intuitively clear that, when we have a clue of the spatial correlation structure underlying the spatial phenomenon to be sampled, it is desirable to exploit this information in the sampling design. In this way we could avoid to duplicate information partly contained in areas already sampled and we can economize sampling costs without loosing in reliability of the estimates. This problem is recognized, for example, by Smartt and Grainger (1974) that, commenting on the various area sampling techniques in vegetation surveys, admit that "...stratification

guards to some extent against any marked clustering of samples..". And later "[..using systematic unaligned sampling..] although some sample clusters may be formed between neighbouring strata, the system is such that this is likely to be rare" (See quotation in Mailing, 1989; p.141).
Furthermore the classical textbook in spatial statistics deal at length with point, transect and quadrat sampling, but there is gap in the definition of *ad hoc* sampling strategies for irregularly shaped data. The simplistic strategy of dealing with these typology of data exactly in the same way as with regular grids of quadrats seems to be misleading in many respects. For instance, the systematic sampling recommended in many textbooks (see Ripley, 1981; Matern, 1960; Cressie, 1993) cannot be employed in a rigorous way when dealing with irregular areas, and the common procedure of superimposing a grid of points on the area to select the locations touched by at least a point leads to absurd designs in presence of elongated shapes and variable tract dimensions as remarked in Arbia (1993).
These reasons should make it clear the need for an explicit introduction of contextual information in spatial sampling designs.

3. The use of context in sampling designs

3.1 Probabilistic designs

The idea of considering the structure of dependency of spatial processes in a probabilistic plan was first considered in the literature by Hedayat et al. (1988). The authors considered the case of time-like observations on a line proposing a technique (termed BSEC after Balanced Sampling design Excluding Continuous units) through which the probability of drawing two neighbouring units is set to zero.
Arbia (1990b, 1991a, 1993) proposes a sampling design (termed DUST after Depending areal Units Sequential Technique) which explicitly considers the spatial correlation properties of the variable to be studied. The basic idea is as follows: "What would you do if you were in a dark room with s candles? You probably would light the first candle in a random point of the room, but then you would find it convenient to light the second candle somewhere further away from the first. How far it depends on the luminosity of the first candle: the stronger the light the further can be located the second candle. You then light the third candle far from the first two and so on. Intensity of the candlelight here plays the role of the degree of spatial correlation." (see Arbia and Switzer, 1994).
The procedure requires an initial estimate of the spatial correlation obtained using a proxy variable or previous surveys. The design then tends to optimality in a sequential way. The first unit is drawn assigning equal probability to all units; the second unit is drawn with a probability that is inversely related to the spatial correlation with the unit already sampled; the third unit is drawn with a probability that is inversely related to the spatial correlation with the two units already sampled and so on. The performances of DUST were tested against simple random design in a series of experiments based on real and simulated data. Espa (1991) shows through simulated data that, when data are positively spatially correlated, DUST outperforms simple random design in terms of efficiency for all sample proportions up to 75%. Arbia (1991b) and Arbia et al. (1991) using real data laid on an irregular grid, show that DUST increases the efficiency of a simple random design of 15%. Furthermore, even if the performances are sensitive to the initial estimate of the spatial correlation, as long as this is of the right sign there is

always an increase in efficiency. Finally Arbia (1992) derives some analytical properties of the DUST estimates and demonstrates formally that the variance of the Horvitz-Thompson estimator of the mean obtained with DUST is smaller than that obtained with a simple random design.

Chandra et al. (1992) defined a Markov sampling scheme based on a sequential drawing of units with variable selection probabilities that depend on the last unit drawn. They use transition probabilities and propose a linear and an exponential specification. In some trivial cases the authors show that the technique is superior to simple random and systematic sampling in terms of the variance of the Horvitz-Thompson estimator of the population total. However the transition probabilities are defined in such a way that it is extremely likely to draw adjacent units due to the short memory of the Markovian process. In fact starting at a random location the second sample unit will tend to be located far from the first one, but the third one will tend to be far from the second one and, hence, again close to the first one.

3.2 Non-probabilistic designs

Non-probabilistic plans look at the combinatorial nature of the problem,. Given N locations and s<N sample locations and having auxiliary information about the underlying spatial correlation, what are the sample locations that minimize a certain objective function amongst the $\binom{N}{n}$ possible samples? Since an exhaustive search is impossible in most relevant real situations, sequential selections or annealing procedures have to be exploited to find sub-optimal solutions. As a consequence a non-probabilistic plan is characterized by two aspects, namely: an optimality criterion, and a search algorithm.

Following this approach in recent years a number of procedures were presented in the literature.

For instance, Sachs and Shiller (1988) suggest to search for the sample locations that minimize the maximum prediction error or, alternatively, that minimize the expected estimation variance. The search for optimality is done by using an annealing algorithm (Geman and Geman, 1984; Van Laarhoven and Aarts, 1987). The authors also report a set of experimental results, but no comparison is made with the results obtained with alternative sampling plans.

Similarly Cressie (1993) suggest to minimize the average prediction variance or, alternatively, minimize the maximum prediction variance. However in the quoted work there is no reference as how to implement the technique and about the performances of these strategies against traditional sampling schemes.

Ferri and Piccioni (1992) minimize a weighted mean of the error variances and apply a modified version of the simulated annealing procedure in order to reach optimality. The procedure is implemented on real data and it is compared with that obtained through a sequential search for optimality, in terms of estimation precision and of CPU time. The annealing procedure outperforms the sequential algorithm. However it should be remarked that the results are data dependent in that no attempt is made to control for the level of spatial correlation and the spatial pattern of variance.

Finally Benedetti and Palma (1993) refer to the concept of “equivalent number of independent observations” (a concept originally due to Clifford and Richardson, 1989 and, independently, to Arbia, 1990a) and to a search procedure to minimize it. In their study they compare the results obtained employing various search algorithms establishing the superiority of the simulating annealing procedure. They also notice lower standard deviations of the estimates of the population mean following this procedure than those obtained with simple random and systematic sample in most experimental cases. However there is no advantage against systematic design when dealing with regular grid data.

All the studies described above, based their conclusions on real and simulated data obtained under very strong hypotheses. First of all, stationarity of the spatial process is invariably assumed. However in practical circumstances spatial data are only seldom stationary through space. For instance agricultural data are characterized by regions displaying the same spatial correlation structure, but this correlation differs dramatically between regions. Secondly, many sets of spatial data present strong heterogeneity of variances and anisotropies as it happens for instance with environmental data (Arbia, 1995) or dealing with surveys where the variance of each observation is related to the size of its location. It is clear that a sampling design which considers the spatial correlation uniform in the area may be rather inefficient (as it is acknowledged e. g. in Arbia, 1991a). On the other hand the knowledge on the variance should be equally incorporated in the design in that one could argue that it is better to economize sample units in zones where the variance is low to spend them where the uncertainty is higher. A third problem not explicitly considered in previous works is that of a stratification prior to sampling.

Arbia and Switzer (1994) proposed a class of designs (termed SCUD after Stratified Correlated units with Unequal variances Designs), and tested their performances by explicitly taking into account the problems described above.

Likewise DUST and other contextual techniques the procedures proposed by Arbia and Switzer (1994) are based on a sequential drawing of locations. There are various arguments in favor to this choice. First of all, sequentiality can help in an applied perspective dealing with repeated surveys where sampling plans are updated through time. Secondly, a sequential drawing can be adopted when one has to reach a given level of precision minimizing the costs, that is the sample size. Thirdly, it is assumed that sequentiality is a rapid way of achieving optimality avoiding the computational burden involved by, for instance, the annealing procedure proposed by Sachs and Shiller (1988) and Ferri and Piccioni (1992). Finally Arbia and Switzer stress that the sequential drawing in the present contest has interesting analogies with the step-wise regression.

Arbia and Switzer (1994) considered a study area partitioned into N non-overlapping locations and the problem of estimating the sample mean using having in hand time series of data at each location. They further considered the estimation variance of the mean as the objective function to be minimized and derived the formal expression for it as a function of the variance covariance matrix between spatial observations. $\mathbf{\Sigma}$. A spatial model for data is then assumed and the matrix $\mathbf{\Sigma}$ of the process is estimated by using temporal observations (reduced to second-order stationarity) at each locations as replications of the spatial field. In particular they considered the data as being generated by an isotropic random field with unknown stationary mean, with variances which may vary spatially, and with a spatial correlation following an exponential declining model. Once the model is specified and $\mathbf{\Sigma}$ is estimated, a draw-by-draw procedure can be activated in which at each step all locations are scanned, the estimation variance is computed and the location which minimizes it is included in the sample.

The model proposed by Arbia and Switzer (1994) was employed by Arbia and Lafratta (1996) to define an optimal sampling plan with reference to pollution data. They considered the practical situation in which data are available at regular intervals (e. g. hourly, daily or weekly) based on sample locations in a study area. Locations are distributed following no rigorous statistical design, but simply based on practical considerations and may or may not be the same in the various surveys. Some or all the sample locations can be (easily) moved from one location to another subject to some minimal constraint. The problem is that of finding an optimal network of monitoring stations at time t given all the information available in the interval [0, t-1] and given a certain objective function. As an objective function they considered the minimum estimation variance for the population mean. They also considered the particular problem, arising in practical circumstances, of evaluating the existing design and updating it by adding, subtracting or moving a limited number of stations from one site to the other. The model was applied to the case of monitoring the air quality in an urban area. The main conclusion is that it is possible to improve precision of the estimates of up to 35% of the estimation variance by simply moving in an optimal way one monitoring station from one place to another.

The model proposed by Arbia and Switzer (1994) and employed in Arbia and Lafratta (1996) is based on the hypothesis that data are generated by an isotropic mean-stationary process. In a subsequent paper Arbia and Lafratta (1997) reconsider the same data set and the same procedure, but relaxed this hypotheses. They achieve this aim by first assuming a non-stationary, non-isotropic process in the geographical space and then transforming it into a non-metric space by employing multidimensional scaling techniques (Sampson and Guttorp, 1992) where isotropy and stationarity can be assumed. They then apply the draw-by-draw sampling technique in the new non-metric space and finally report the observations in the original space. Also in this case the technique proved useful in the reduction of the estimation variance of the sample mean estimator.

Further generalizations of Arbia-Switzer sampling strategy are currently under study where a more general definition of spatial dependence is considered (wider than linear correlation) and where interested is restricted in the tails of the bivariate margins to study extreme events (like pollution threshold trespassing in environmental studies or flood or other agricultural disasters) (see Arbia, 2001a, 2001b).

5. Conclusions

This paper reviews spatial sampling designs in the presence of spatial dependence. We discuss the use of traditional sampling strategies (such as random or systematic) in the spatial context and we motivate the need for *ad hoc* techniques for a better exploitation of the information content of the sample and for achieving higher levels of efficiency.

We referred to the techniques that exploit contextual information as “contextual designs”. When dealing with contextual design a further distinction has been made between probabilistic and non probabilistic designs. In the first instance (likewise in random sampling) we assign a certain drawing probability to each spatial unit. In the second instance (likewise in systematic sampling) once the sample size is given and an objective function is specified, the sampling plan is fixed.

In the paper we review some of the techniques proposed in recent years following both the probabilistic and the non probabilistic strategy, and we report some result of comparison studies with the traditional non contextual solutions.
We also argue in favour of a draw-by-draw strategy that appears to be a convenient way of achieving optimality by reducing the computational burden, can be employed to update previously defined designs and can help in defining rigorous systematic sampling when areal data refer to irregularly shaped locations.
A secondary aim of the paper was that of providing a (tentatively) exhaustive bibliography on the topic.
Despite the wide literature examined there is no clear indication on what is the best sampling strategy to be adopted in practical instances. One clear conclusion seems to be the superiority of contextual plans in terms of the efficiency that can be achieved. However there is no clear evidence that a plan which is optimal for regular lattices, is also a good one when dealing with irregular lattices and further work is required in this direction.
Further topics that deserve more attentions are: the specification of multiple variables objective functions; the definition of optimal sequential plans to achieve a given level of accuracy with the minimum sample size; the extension to three dimensions; the effects of sample design on features other than the sample mean. Related to this last point it would be interesting, within an applied perspective, to define designs with the explicit objective of locating extremes in the plane.

References

Arbia G. (1990a) On second order non-stationarity in two-dimensional lattice processes, *Computational Statistics and Data Analysis*, 9, 147-160,.

Arbia G. (1990b) Sampling dependent spatial units, Paper presented at the Workshop on spatial statistics, Commission on mathematical modelling of the IGU, Boston.

Arbia G. (1991a) "GIS-based sampling design procedures", in J. Harts, H. Offens and H. Schotten (eds.) *Proceedings of EGIS '91*, *Second European Conference on Geographical Information Systems*, Brussels, EGIS Foundation, Vol. 1, Utrecht, pp. 27-35.

Arbia G. (1991b) L'impiego dei sistemi geografici informativi nelle rilevazioni ufficiali, Ann. di Stat., ISTAT, Roma.

Arbia G. (1992) Campionamento areale in presenza di correlazione spaziale, Proceedings of the XXXVI meeting of the Italian Statistical Society, 1, 311-21.

Arbia G. (1993) The use of GIS in spatial surveys*, International Statistical Review*, 61, 2, 339-59.

Arbia, G. (1995) Updating existing sampling designs in repeated environmental surveys, *Working paper,* 95.6, Department of statistical sciences, University of Padua, Italy.

Arbia, G. (2001a) Non-linear spatial dependence in the tails", Paper to be presented at the *Royal Statistical Society theme conference on SPATIAL MODELLING, 2001, Glasgow 3-6 July 2001*

Arbia, G. (2001b) Exploring non-linear spatial dependence in the tails of the bivariate marginals of a random field, Paper presented at *the Applied Econometric*

Association (AEA) Conference held in conjunction with WRSA Special Fortieth Anniversary Annual Meeting, Palm Springs, California ; February 25 – 28, 2001.

Arbia G., and Switzer P. (1994) Spatial sampling designs for stratified correlated units with unequal variances, Working Paper, n. 94.6, Department of Statistical Sciences, Faculty of Statistics, Padua.

Arbia G., Espa G. and Genovese G. (1991) L'organizzazione di indagini campionarie tramite Sistemi Informativi Territoriali, Proceedings of the XII meeting of the Italian Regional Sciences Association, 2, 147-68.

Arbia, G. and Lafratta, G. (1996) Evaluating and updating the sample designs in repated environmental surveys: monitoring air quality in Padua, *Journal of Agricultural, biological and environmental statistics, 2, 4, 451-466.*

Arbia, G. and Lafratta, G. (1997) Spatial sampling designs optimised under anisotropic superpopulation models, Quaderni di statistica, University of Chieti (submitted for publication).

Barnes R. J. (1989) A partial history of spatial sampling design, Geostatistics, 10-13.

Benedetti R. and Palma D. (1993) A GIS-based spatial units selection method, Proceedings of EGIS '93, Genova.

Bras R. L. and Colon R. (1978) Time averaged areal mean precipitation: estimation and network design. Water Resourches Research, 14, 5, 878-88.

Bras R. L. and Rodriguez-Iturbe I (1976) Rainfall network design for runoff prediction, Water resourches Research, 12, 6, 1197-1208.

Chandra K. S., Sampath A. and Balasubramani G. K. (1992) Markov sampling for finite populations, *Biometrika*, 79, 210-3.

Cliff A. D. and Ord J. K. (1981) *Spatial Processes: Models and Applications*, Pion, London.

Clifford P. and Richardson S. (1989) Testing the association between two spatial processes, Statistics & Decision, Supplement, 2, 155-60.

Cochran W. G. (1946) Realtive accuracy of systematic and stratified samples for a certain class of populations, Annals of Mathematical Statistics,164-177.

Cochran W.G. (1963) Sampling techniques, Wiley.

Cressie N. (1993) *Statistics for Spatial Data* (revised edition), John Wiley & Sons, New York.

Dalenius T. Hajek J. and Zubrzycki S. (1961) On plane sampling and related geometrical problems. Proceedings of the 4th Berkley Symposium on Math. Stat. and Probability, Berkley, 1, 125-150.

Das A. C. (1950) Two dimensional systematic sampling and the associated stratified and random sampling, Sankhya, 10, 95-108.

Espa G. (1991) Metodi di campionamento per dati areali dipendenti, Contributi di Statistica Economica, 3, Dipartimento di Contabilita' Nazionale, Universita' La Sapienza, Roma,

Ferri M and Piccioni M. (1992) Optimal selection of statistical units: an approach via simulated annealing, Computational statistics and data analysis, 13, 47-62.

Finney D. J. (1948) Random and systematic sampling in timber surveys, Forestry, 22, 64-99.

Finney D. J. (1950) An example ogf periodic variation in forest sampling, Forestry, 23, 96-111.

Finney D. J. (1953) The estimation of error in systematic sampling of forests, Journal of Indian Society of Agricultural Statistics, 5, 6-16.

Geman S. & Geman D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE, Transaction on pattern analysis and machine intelligence, Vol. PAMI-6, n.6.

Ghosh B. (1949) Topographic variation in statistical fields, Calcutta Statistical Association Bullettin, 2, 11-28.

Hansen M.H., Hurwitz W.N. e Madow W.G. (1960) Sample survey. Methods and theory, Wiley, New York.

Hasel A. A. (1938) Sampling error in timber surveys, Journal of Agricultural Research, 57, 713-736.

Hedayat A.S., Rao C.R. and Stufken J. (1988) Sampling plans excluding contiguous units, Journal of Statistical Planning and Inference, 19, 159-17

Holmes J. H. (1970) The theory of plane sampling and its application in geographic research, Economic Geography, 46, 379-91.

Kish L. (1965) Survey sampling, Wiley, New York.

Krige D. G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand, Journal of Chem. Metall. Mining Society S. Africa, 52, 119-32.

Laarhoven Van P. J. M. and Aarts E. H. L. (1987) Simulated annealing theory and practice, Reidel, Dordrecht.

Mailing D.H.(1989) Measurements from maps, Pergamon Press.

Matern B. (1947) Metoder att uppskatta noggrannheten vid linje- och provytetaxering, Medd. Fran statens skogsforskningsints, 36, 1.

Matern B. (1960) Spatial variation, Meddelanden fran Statens Skogsforskningsinstitut, 49, No. 5 [Lecture notes in statistics, n. 36, Springer Verlag, New York (1960) (reprint)].

Matern B. (1986) Spatial variation, second edition, Springer-Verlg, Berlin.

Matheron G. (1962) Traite' de geostatistique, Vol. 1, Editions Technip, Paris.

Matheron G. (1963) Traite' de geostatistique, Vol.2 , Editions Technip, Paris.

Matheron G. (1965) Les variables regionalisees et leur estimation, Editions Masson, Paris.

Milne A. (1959) The centric systematic area-sample treated as a randon sample, Biometrics, 15, 270-97.

Payandeh B. (1970) Relative effiecency of two-dimensional systematic sampling, Forestry Science, 16, 312-17.

Quenouille M.H. (1949) Problems in plane sampling, Ann. of Math. Stat., 20, 355-75.

Ripley B. D. (1981) Spatial statistics, Wiley, New York.

Rodiriguez-Iturbe I and Mejia J. M.,(1974) The design of rainfall networks in time and space, New York.

Sacks J. and Shiller S. (1988) Spatial designs, in *Statistical decision theory and related topics*, Vol. 2, Gupta and Berger Eds., Springer Verlag, 385-399.

Sampson P. D. and Guttorp, P. D. (1992) Nonparametric estimation for monitoring data, in *Statistics in environmental and earth sciences,* eds. Guttorp, P. and Walden, A., London: Griffin, 52-70.

Smartt P.F.M. and Grainger J.E.A. (1974) Sampling for vegetation survey, Journal of Biogeography, 1, 193-206.

Stephan F. (1934) Sampling errors and interpretations of social data ordered in time and space, Proceedings of the Am. Stat. Journal, New series, 185A, Journal of the American Statistical Association, 29, suppl., 465-6.

Viera S. R. , Nielsen D. R. and Biggar J. W. (1981) Spatial variability of field-measures infiltration rates, Soil Sciences Soc. Am. Journal, 45, 1040-8.

Water Resourches Research (1979) n. 15, 6.

Webster R. and Oliver M.A. (1990) Statistical methods in soil and land resourches survey, Oxford Univ. Press.

Whittle P. (1952) Tests of fit in time series, Biometrika, 39.

Whittle P. (1963) Stochastic processes in several dimensions, Bull. International Statistical Institute, 40, 974-94.

Williams R. M. (1956) The variance f the mean of systematic samples, Biometrika, 43, 137-48.

Yaglom A. M. (1957) Some classes of random fields in n-dimensional space, related to stationary random processes, Theor. Probability Appl., 2, 273-320.

Yaglom A. M. (1961) Second-order homogeneous random fields, Proceedings of the 4th Berkeley Symposium on Math, Stat. and Probability, University of California Press, Berkeley, 2, 593-622.

Yaglom A. M. (1962) An introduction to the theory of stationary random functions, Prenctice-Hall, Englewood Cliffs, New Jersey.

Zubrzycky S. (1958) Remarks on random stratified and systematic sampling in a plane, Colloquium Mathematicum, 6, 251-64.

Small Area Estimation with Applications to Agriculture

J.N.K. Rao
Carleton University, School of Mathematics and Statistics
Ottawa, Canada, e-mail: jrao@math.carleton.ca

Abstract: Small area estimation has received a lot of attention in recent years due to growing demand for reliable small area statistics that are needed for formulating policies and programs. Indirect estimates that "borrow strength" from areas are used because direct area-specific estimates may not be reliable due to small area-specific sample sizes. This paper provides an appraisal of indirect estimates, both traditional and model-based. Several applications to agriculture are also presented.

Key words and phrases: Area level model, composite estimate, unit level model.

1. Introduction

In the context of agriculture, the term "small area" generally refers to a small geographical area, such as a "tehsil" or block in India. Reliable small area information on crop statistics is needed for formulating agricultural policies.

Crop yield statistics are generally obtained through sample surveys. For example, crop cutting experiments in sampled fields are used in India to obtain direct (or area-specific) estimates of crop yields. Data collected from sample surveys can be used to derive reliable direct estimates for large areas, such as a district, making effective use of auxiliary data. For example, in India remote sensing satellite data are currently used as auxiliary information to produce reliable direct estimates of crop areas and crop yields at the district level. Singh and Goel (2000) proposed poststratification of the crop area on the basis of Vegetation Indices derived from the satellite data. The post-stratified estimators were considerably more efficient than the traditional estimators using only geographical stratification.

Direct area-specific estimates may not provide acceptable precision at the small area level because sample sizes in small areas are seldom large enough. In some cases, sample sizes at a higher level (such as a state) also may be small (cf., example 3.2, Section 3). This makes it necessary to "borrow strength" from related areas to find indirect estimates that increase the effective sample size and thus increase the precision. Such indirect estimates are based on either implicit or explicit models that provide a link to related small areas through supplementary data such as data from a recent census of agriculture, remote sensing satellite data and administrative records.

Indirect estimates based on implicit models include synthetic and composite estimates. Indirect estimates based on explicit models have received a lot of attention because of the following advantages over traditional synthetic and composite estimates: (1) Model-based methods make specific allowance for local variation through complex error structures in the model that links the small areas. (2) Models can be validated from the sample data. (3) Methods can handle complex cases such as cross-sectional and time series data. (4) Stable area specific measures of variability associated with the estimates may be obtained, unlike overall measures commonly used for traditional indirect estimates.

In this paper, we provide a brief account of small area estimation in the context of agricultural surveys.

2. Design Issues

Efficient methods of designing surveys for use with direct estimates of large area totals or means received a lot of attention over the past 50 years or so. But surveys design issues that have an impact on small area statistics should also be considered. Singh, Gambino and Mantel (1994) proposed several methods for use at the design stage to minimize the use of indirect small area estimates. Those methods include (i) replacing clusters by using list frames, (ii) use of many strata to provide better sample size control at the small area level and (iii) compromise sample allocation. They presented an excellent illustration of compromise sample size allocation in the Canadian Labour Force Survey to satisfy reliability requirements at the provincial level as well as subprovincial level. Preventive measures, such as (i)-(iii), should be undertaken at the design stage, whenever possible, to achieve adequate precision using direct estimates. Other methods for use at the design stage include the use of two or more (possibly) incomplete frames, combining data from rolling samples and integration of surveys.

Preventive measures at the design stage may minimize the need for indirect estimates significantly, but for some small areas (e.g., tehsils in India) sample sizes may not be large enough to provide adequate precision using direct estimates even after implementing such measures.

3. Synthetic and Composite Estimates

Suppose the population is divided into g large post-strata for which reliable direct estimates of the poststrata totals, $Y_{.g}$, can be calculated from the survey data, where $Y_{.g} = \Sigma_i Y_{ig}$ and Y_{ig} is the total of the characteristic of interest, y, for the units in small area i that belong to post-stratum g. Our interest is to estimate the small area totals $Y_i = \Sigma_g Y_{ig}$, $i = 1,\ldots,m$ using known auxiliary totals X_{ig}.

3.1 Synthetic estimates

A synthetic estimate of Y_i is given by

$$\hat{Y}_i^S = \sum_g X_{ig} \left(\hat{Y}_{.g} / \hat{X}_{.g} \right) \tag{3.1}$$

where $\hat{Y}_{.g}$ and $\hat{X}_{.g}$ are reliable direct estimates of poststrata totals $Y_{.g}$ and $X_{.g}$. The design-bias of $\hat{Y}_i^S$ under repeated sampling will be small relative to the total Y_i if the ratios $R_{ig} = Y_{ig} / X_{ig}$ are homogeneous across small areas, i.e., $R_{ig} \equiv R_{.g} = Y_{.g} / X_{.g}$ for each g. Moreover, the design standard error of $\hat{Y}_i^S$ will be small relative to Y_i since it depends only on the variances and covariances of poststrata estimates $\hat{R}_{.g} = \hat{Y}_{.g} / \hat{X}_{.g}$. Thus, the synthetic estimate $\hat{Y}_i^S$ will be reliable under the assumption $R_{ig} \equiv R_{.g}$. But such an assumption may be quite strong in practice, and in fact $\hat{Y}_i^S$ can be heavily biased for small areas exhibiting strong individual effects.

The variance of $\hat{Y}_i^S$ is readily estimated, but it is more difficult to estimate the mean squared error (MSE) of $\hat{Y}_i^S$. An approximately unbiased estimate of MSE ($\hat{Y}_i^S$) is given by

$$\text{mse}\left(\hat{Y}_i^S\right) = \left(\hat{Y}_i^S - \hat{Y}_i\right)^2 - v\left(\hat{Y}_i\right) \qquad (3.2)$$

where $\hat{Y}_i$ is the direct estimate of Y_i and $v\left(\hat{Y}_i\right)$ is a design-unbiased estimate of variance of $\hat{Y}_i$. However, the MSE estimate (3.2) may be very unstable. Consequently, it is a common practice to average them over i to get a stable estimate of MSE. But such a global measure of variability that does not vary over areas can be misleading in practice.

Example 3.1
Singh and Goel (2000) used a synthetic estimate of the form (3.1) to estimate crop yields in India at the tehsil level. Poststrata were formed on the basis of Vegetation Indices derived from the remote sensing satellite data. The survey estimates $\hat{Y}_{.g}$ were obtained from crop yield surveys based on crop cutting experiments. Actually, Singh and Goel (2000) used the totals $X_{.g}$ instead of $\hat{X}_{.g}$ in (3.1) where $X_{.g}$ is the total crop area in $g-$th poststratum. On the basis of standard error of $\hat{Y}_i^S$, their evaluation study indicated that $\hat{Y}_i^S$ is often significantly more efficient than $\hat{Y}_i$ at the tehsil level and especially at the block level. However, ignoring the bias and using the standard error may be overly optimistic.

3.2 Composite estimates

A simple way to balance the potential bias of synthetic estimate, $\hat{Y}_i^S$, against the instability of the direct estimate, $\hat{Y}_i$, is to take a weighted average of the two estimates. This leads to a composite estimate of the form

$$\hat{Y}_i^C = \alpha_i \hat{Y}_i + (1 - \alpha_i)\hat{Y}_i^S \qquad (3.3)$$

for some suitably chosen weight a_i in the range [0,1]. Optimal weights that minimize MSE$\left(\hat{Y}_i^S\right)$ can be obtained, but their estimates, $\hat{a}_i$, can be very unstable as they involve mse$\left(\hat{Y}_i^S\right)$ given by (3.2). Purcell and Kish (1980) used a common weight, a, and then minimized the average MSE over small areas. This leads to a weight $\hat{a}$ of the form $1 - m\bar{v} / \sum_i \left(\hat{Y}_i^S - \hat{Y}_i\right)^2$, where $\bar{v} = m^{-1} \sum_i v\left(\hat{Y}_i\right)$ and m is the number of small areas. But

the use of a common weight may not be reasonable if the individual variances, $V(\hat{Y}_i)$, vary considerably.
Formula (3.2) with $\hat{Y}_i^S$ replaced by $\hat{Y}_i^C$ is often used to estimate $\text{MSE}(\hat{Y}_i^C)$. But this MSE estimate is also very unstable.

Example 3.2
Eklund (1998) used a composite estimate of the form (3.3) to estimate net coverage error for the 1997 U.S. Census of Agriculture at the state (small area) level. Survey data were used to estimate net coverage error. But the sample size at the state level were not large enough for the direct estimates at state level to be reliable, unlike the direct estimates at the region level. Eklund used a synthetic state estimate of the form (3.1) without post-stratification. It is given by

$$\hat{Y}_i^S = (X_i / X_R)\hat{Y}_R,$$

where $\hat{Y}_R$ is a direct estimate of the regional total Y_R , X_i and X_R are the census totals for state i and region R of the corresponding characteristic of interest. The synthetic estimate $\hat{Y}_i^S$ was combined with the direct estimate $\hat{Y}_i$ using a composite estimate $\hat{Y}_i^C$.
Because of the instability in the MSE estimates of the form (3.2), $\text{mse}(\hat{Y}_i^C)$, Eklund proposed to smooth the estimates by modelling the relative mean square error, $\text{mse}(\hat{Y}_i^C)/(\hat{Y}_i^C)^2$, using linear regression on the census state total X_i. Eklund noted some difficulties with this method.

4. Area Level Models

We now turn to model-based methods based on small area linking models involving random small area effects. Such models may be broadly classified into two types: (a) area level models; (b) unit level models. In this section we study area level models and delegate unit level models to section 5.
A basic area level model that uses area level covariates has two components: (a) Direct survey estimate $\bar{y}_i$ of the $i-$th area mean $\bar{Y}_i$, possibly transformed as $\hat{\theta}_i = g(\bar{y}_i)$, is equal to the sum of the population value $\theta_i = g(\bar{Y}_i)$ and the sampling error e_i:

$$\hat{\theta}_i = \theta_i + e_i, \quad i = 1,...,m, \tag{4.1}$$

where the e_i's are assumed to be independent across areas with means 0 and known variances ψ_i. (b) A linking model that relates the θ_i's to area level covariates $\mathbf{z}_i = (z_{1i},...,z_{pi})^T$ through a linear regression model:

$$\theta_i = z_i^T \boldsymbol{\beta} + v_i, \quad i = 1,\ldots,m, \tag{4.2}$$

where the model errors v_i are assumed to be independent and identically distributed with mean 0 and variance σ_v^2. The parameter σ_v^2 is a measure of homogeneity of the areas after accounting for the covariates $\mathbf{z}_i$. Combining (4.1) and (4.2), we get a mixed linear model

$$\hat{\theta}_i = z_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1,\ldots,m. \tag{4.3}$$

Using the data $\{(\hat{\theta}_i, \mathbf{z}_i), \ i = 1,\ldots,m\}$ we can obtain estimates, θ_i^*, of the realized values of θ_i from the model (4.3). A model-based estimate of $\overline{Y}_i$ is then given by $g^{-1}(\theta_i^*)$. Note that the model involves both design-based random variables, e_i, and model-based random variables, v_i.

Empirical beat linear unbiased prediction (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB) methods have played a prominent role in the estimation of $\overline{Y}_i$ under model (4.3). EBLUP method is applicable for mixed linear models and EBLUP estimates do not require normality assumption on the random errors v_i and e_i. On the other hand, EB and HB are more generally applicable under specified distributional assumptions. HB methods lead to exact inferences via posterior distributions: $p(\theta_i \mid \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1,\ldots,\hat{\theta}_m)^T$. EBLUP and EB estimate of θ_i are identical under normality and nearly equal to the HB estimate $E(\theta_i \mid \hat{\boldsymbol{\theta}})$, but measures of variability of the estimates may differ. Under EBLUP and EB we use an estimate of $\mathrm{MSE}(\tilde{\theta}_i) = E(\tilde{\theta}_i - \theta_i)^2$ as a measure of variability of $\tilde{\theta}_i$, where the expectation is with respect to the model (4.3). On the other hand, HB uses posterior variance $V(\theta_i \mid \tilde{\boldsymbol{\theta}})$ as a measure of variability.

EBLUP estimate of θ_i is a composite estimate of the form

$$\theta_i^* = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \mathbf{z}_i^T \hat{\boldsymbol{\beta}}, \tag{4.4}$$

where $\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \psi_i)$ and $\hat{\boldsymbol{\beta}}$ is the weighted least squares estimate of $\boldsymbol{\beta}$ with weights $(\hat{\sigma}_v^2 + \psi_i)^{-1}$ obtained by regressing θ_i on $\mathbf{z}_i$: $\hat{\boldsymbol{\beta}} = \left(\sum_i \hat{\gamma}_i \mathbf{z}_i \mathbf{z}_i^T\right)^{-1} \left(\sum_i \hat{\gamma}_i \mathbf{z}_i \theta_i\right)$ and $\hat{\sigma}_v^2$ is an estimate of the variance component σ_v^2. That is, the EBLUP estimate, θ_i^*, is a weighted combination of the direct estimate, $\hat{\theta}_i$, and a regression synthetic estimate $\mathbf{z}_i^T \hat{\boldsymbol{\beta}}$ with weights $\hat{\gamma}_i$ and $1 - \hat{\gamma}_i$ respectively. The EBLUP estimate gives more weight to the direct estimate when the sampling variance, ψ_i, is small (or $\hat{\sigma}_v^2$ is large) and moves towards the regression synthetic estimate as ψ_i increases (or $\hat{\sigma}_v^2$ decreases).

For the nonsampled areas, the EBLUP estimate is given by the regression synthetic estimate, $\mathbf{z}_i^T \hat{\boldsymbol{\beta}}$, using the known covariates associated with the nonsampled areas.

A lot of attention has been given to the estimation of MSE or the posterior variance, $V(\theta_i \mid \tilde{\boldsymbol{\theta}})$, under the HB set-up. Complex models under HB can be handled using recently developed Monte Carlo Markov Chain (MCMC) methods. We refer the reader to Ghosh and Rao (1994) and Rao (1999) for some recent developments on the estimation of MSE and the computation of posterior variance.

Under model (4.4), the leading term of $\text{MSE}(\tilde{\theta}_i)$ is given by $\gamma_i \psi_i$ which shows that the EBLUP estimate can lead to large gains in efficiency over the direct estimate with variance ψ_i, when γ_i is small (or the model variance, σ_v^2, is small relative to the sampling variance, ψ_i). The success of small area estimation, therefore, largely depends on getting good auxiliary data $\{\mathbf{z}_i\}$ that can lead to a small model variance relative to sampling variance. One should also make a thorough internal validation of the assumed model.

Sampling variances, ψ_i, may not be known in practice in which case one often resorts to smoothing of the estimated design variances, $\hat{\psi}_i$, to get stable estimates ψ_i^*, say. The smoothed estimate ψ_i^* is then treated as a proxy to ψ_i.

Example 4.1

Fuller (1981) applied the area level model (4.3) to estimate mean soybean hectares per segment in 1978 at the county level. He used the mean number of pixels of soybeans per area segment, z_{2i}, ascertained by satellite imagery and the mean soybean hectares from the 1974 U.S. Agricultural Census, z_{3i}, as county (area) level covariates. Survey estimates, $\bar{y}_i$, for a sample of $m = 10$ counties were obtained by sampling area segments within sampled counties.

Fuller obtained model-based estimates of the population means, $\bar{Y}_i$, for the sampled counties as well as the nonsampled counties. His model is given by

$$\bar{y}_i - z_{3i} = \beta_0 + \beta_1 z_{2i} + \beta_2 z_{3i} + v_i + e_i \tag{4.5}$$

with known error variances $\sigma_v^2 = 25$ and $\sigma_e^2 = 18$. Note that z_{2i} and z_{3i} are known for all the 16 counties. The model (4.5) is a special case of (4.3) with $\hat{\theta}_i = \bar{y}_i - z_{3i}$ and $\psi_i = \psi = \sigma_e^2$. Fuller's estimate of $\bar{Y}_i$ for sampled counties is obtained from (4.4) as

$$\bar{y}_{iF} = g^{-1}(\theta_i^*) = \theta_i^* + z_{3i} = z_{3i} + \mathbf{z}_i^T \hat{\beta} + \gamma(\bar{y}_i - z_{3i} - \mathbf{z}_i^T \hat{\beta}), \quad i = 1,\ldots,10, \tag{4.6}$$

where $\gamma = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2) = 25/43$.

For the nonsampled counties, $\bar{y}_{iF}^* = z_{3i} + \mathbf{z}_i^T \hat{\beta}, \quad i = 11,\ldots,16$.

Under model (4.5), Fuller calculated the total MSE of the estimates $\bar{y}_{iF}$ for the sampled counties as 127 and for the nonsampled counties as 210. The total MSE for all the 16 counties equals 127+210 = 337. On the other hand, if one were to use the prior census mean z_{3i} as the estimate of $\bar{Y}_i$, then the total MSE of z_{3i} under model (4.5) is estimated as 1287 which is considerably larger than 337. The model-based estimates, $\bar{y}_{iF}$, outperformed the prior census predictors, z_{3i}, in terms of total MSE. They are also better than the direct estimates $\bar{y}_i$ in terms of total MSE for the sampled counties: Total MSE of $\bar{y}_i$'s =10 σ_e^2=180 compared to 127, the total MSE of $\bar{y}_{iF}$'s.

5. Unit Level Models

A basic unit level model assumes that the unit y-values, y_{ij}, associated with the $j-$th population unit $(j=1,...,N_i)$ in the $i-$th area is related to unit-level covariates, $\mathbf{x}_{ij}$, for which the population mean vector $\bar{\mathbf{X}}_i$ is known. If y is a continuous response (e.g., crop yield), we assume a one-fold nested error linear regression model

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad j = 1,\ldots,N_i; \;\; i = 1,\ldots,m \tag{5.1}$$

where the random sample area effects v_i have mean 0 and common variance σ_v^2 and independently distributed. Further, the v_i are independent of the residual errors e_{ij} which are assumed to be independently distributed with mean 0 and common variance σ_e^2 (Battese, Harter and Fuller, 1988).

If N_i is large, the population mean $\bar{Y}_i$ is approximately equal to $\bar{\mathbf{X}}_i^T\beta + v_i$. The sample data $\{y_{ij}, \mathbf{x}_{ij}, \; j=1,...,n_i; \; i=1,...,m\}$ is assumed to obey the population model (5.1). This implies that sample selection bias is absent which is satisfied by simple random sampling within areas. For more general sampling designs, the sample data will satisfy the assumption if the selection probabilities, p_{ij}, depend only on the auxiliary variables in $\mathbf{x}_{ij}$; for example, probability proportional to size (PPS) sample, where size is used as an auxiliary variable in the model (5.1). Nonprobability samples obeying the model (5.1) can also be used to estimate the mean $\bar{Y}_i$ (Example 5.1).

We assume that $\bar{Y}_i = \bar{X}_i^T\beta + v_i$. Then the EBLUP estimate of $\bar{Y}_i$ is a composite estimate of the form

$$\bar{y}_i^* = \hat{\gamma}_i \left[\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T \hat{\boldsymbol{\beta}} \right] + (1-\hat{\gamma}_i)\bar{X}_i^T \hat{\boldsymbol{\beta}}, \;\; i = 1,...,m \tag{5.2}$$

where $\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})$ with estimated variance components $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$, and $\hat{\boldsymbol{\beta}}$ is the weighted least squares estimate of $\boldsymbol{\beta}$ which depends on $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$. As the small area sample size n_i increases, the EBLUP estimate approaches the "survey regression" estimate $\bar{y}_i + (\overline{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T \hat{\boldsymbol{\beta}}$. On the other hand, for small n_i and small $\hat{\sigma}_v^2 / \hat{\sigma}_e^2$ it tends towards the regression synthetic estimate $\overline{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}$. For the nonsampled areas, $\bar{y}_i^* = \overline{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}$. Note that the EBLUP estimates $\bar{y}_i^*$ do not depend on the survey weights, w_{ij}, so that design-consistency as n_i increases is forsaken except when the design is self-weighting as in the case of simple random sampling. On the other hand, the EBLUP estimate under the area level model (4.3) is design-consistent because the direct estimate $\hat{\theta}_i$, used in (4.4), is design-consistent.

The variance components σ_v^2 and σ_e^2 are estimated from the sample data, using the well-known method of fitting constants (Battese, Harter and Fuller, 1988) or the restricted maximum likelihood (REML) method (assuming normality of the errors v_i and e_{ij}).

Under model (5.1) for the sample data, the leading term of $\text{MSE}(\bar{y}_i^*)$ is given by $\gamma_i (\sigma_e^2 / n_i)$ which shows that the EBLUP estimate can lead to large gains in efficiency over the survey regression estimate when γ_i is small. Note that the leading term of MSE of the survey regression estimate equals σ_e^2 / n_i.

It is possible to incorporate survey weights under unit level models, using model-assisted estimates (Prasad and Rao, 1999). Further research on this topic would be practically useful.

Example 5.1

Battese, Harter and Fuller (1988) applied the nested error regression model to estimate area under corn and soybeans for each of $m = 12$ counties in North-central Iowa, using farm interview data in conjunction with LANDSAT satellite data. Each county was divided into area segments, and the areas under corn and soybeans were ascertained for a sample of segments by interviewing farm operators; the number of sample segments, n_i, in a county ranged from 1 to 5. Unit level auxiliary data in the form of number of pixels classified as corn and soybeans were also obtained for all the area segments, including the sample segments, in each county using the LANDSAT satellite readings. In this application, $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})^T$ where x_{1ij} and x_{2ij} respectively denote the number of pixels classified as corn and the number of pixels classified as soybeans in the $j-$th area segment of the $i-$th county, and y_{ij} denotes the number of hectares of corn (soybeans) in the $j-$th sample area segment of the $i-$th county.

The sample data for the second sample segment in Hardin county ($n_i = 2$) was deleted because the corn area for that segment looked erroneous in a preliminary analysis. Battese, Harter and Fuller calculated the ratio of the model-based standard error of the EBLUP estimate to that of the survey regression estimate. This ratio decreased from about 0.97 to 0.77 as the number of sample segments, n_i, decreases from 5 to 1. The reduction in standard error is considerable when $n_i \leq 3$.

The EBLUP estimates were adjusted to agree with the reliable survey regression estimate for the entire area covering the 12 counties. This adjustment produced a very small increase in the standard errors.

Battese, Harter and Fuller (1988) also reported some methods for validating the assumed model. First, they introduced quadratic terms x_{1ij}^2 and x_{2ij}^2 into the model and tested the null hypothesis that the regression coefficients associated with the quadratic terms are zero. The null hypothesis was not rejected at the 5% level. Secondly, they tested the null hypothesis that the error terms v_i and e_{ij} are normally distributed by using the transformed residuals $(y_{ij} - \hat{\alpha}_i \bar{y}_i) - (\mathbf{x}_{ij} - \hat{\alpha}_i \bar{x}_i)^T \hat{\boldsymbol{\beta}}$ with $\hat{\alpha}_i = 1 - (1 - \hat{\gamma}_i)^{1/2}$. Under the null hypothesis, the transformed residuals are independent normal with mean 0 and variance σ_e^2. The well-known Shapiro-Wilk W statistic, applied to the transformed residuals, gave P-values equal to 0.921 and 0.299 for corn and soybeans respectively, suggesting the tenability of the normality assumption.
Recently, I have used the HB approach and calculated "posterior predictive" probabilities to test the model fit. Under this criterion, extreme probabilities (closer to 0 or 1) suggest poor fit and values closer to 0.5 indicate good fit. I obtained values equal to 0.507 for corn and 0.503 for soybeans, suggesting very good fit of the Battese-Harter-Fuller model with $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})^T$.

Example 5.2
Stasny, Goel and Ramsey (1991) used a regression synthetic estimate to produce county estimates of wheat production in Kansas, USA. They used a non-probability sample of farms, assuming a linear regression model (without the small area effect v_i) relating wheat production of $j-$th farm in $i-$th county to predictor variables with known county totals. A measure of farm size was included as a predictor variable to account for the fact that the sample was not a probability sample. A ratio adjustment to the synthetic estimates was made to ensure that the adjusted estimates add up to the direct state estimates of wheat production obtained from a large probability sample.

6. Discussion

Preventive measures at the design stage, such as those outlined in Section 2, may reduce the need for indirect estimates significantly. But, for many applications sample sizes in many small areas may not be large enough to achieve adequate precision through direct estimates even after taking such measures.
We have presented model-based small area estimation under a basic area level model and a basic unit level model. Various extensions of the basic models have been studied, including multivariate and time series models and logistic mixed linear models for binary responses (cf., Rao, 1999).
Good auxiliary information related to variables of interest, such as remote sensing satellite data related to crop yields, plays a vital role in model-based estimation.
Model validation also plays an important role.
Indirect estimates of small area totals or means, such as the EBLUP estimates, may not be suitable if the objective is to identify areas with extreme population values or to rank areas or to identify areas that fall below or above certain specified level. Ghosh and Rao (1994) and Louis (2001) reviewed some methods for handling such cases.

Finally we should emphasize the need for formulating an overall program that covers issues related to sample design and data development, organization and dissemination, in addition to those pertaining to methods of estimation for small areas.

References

Battese, G.E., Harter, R.M., Fuller, W.A. (1988) An error-components model for prediction of crop areas using survey and satellite data, *Journal of the American Statistical Association*, 83, 28-36.

Eklund, B. (1998) Small area estimation of coverage error for the 1997 Census of Agriculture, in: *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 335-338.

Fuller, W.A. (1981) Regression estimation for small areas, in: *Rural America in Passage: Statistics for Policy,* National Academy Press, Washington, D.C., 572-586.

Ghosh, M., Rao, J.N.K. (1994) Small area estimation: an appraisal, *Statistical Science*, 9, 55-93.

Louis, T.A. (2001) Bayes/EB ranking, histogram and parameter estimation, in: *Empirical Bayes and Likelihood Inference*, Ahmed, S.E. and Reid, N. (Eds.), Springer, 1-16.

Prasad, N.G.N., Rao, J.N.K. (1990) The estimation of the mean squared error of small-area estimators, *Journal of the American Statistical Association*, 85, 163-171.

Purcell, N.J., Kish, L. (1980) Postcensal estimates for local areas, *Biometrics*, 35, 365-384.

Rao, J.N.K. (1999) Some recent advances in model-based small area estimation, *Survey Methodology*, 25, 175-186.

Singh, M.P., Gambino, J., Mantel, H.J. (1994) Issues and strategies for small area data, *Survey Methodology*, 20, 3-22.

Singh, R., Goel, R.C. (2000) Use of remote sensing satellite data in crop surveys, *Technical Report, Indian Agricultural Statistics Research Institute*, New Delhi.

Stasny, E.A., Goel, P.K., Ramsey, D.J. (1991) County estimates of wheat production, *Survey Methodology*, 17, 211-225.

Sample Designs for National Surveys: Surveying Small-Scale Economic Units

Vijay Verma
ORC Macro International Social Research, London
Angel Corner House, 1 Islington High Street, London N1 9AH United Kingdom.
tel (44-20) 7675 1063; fax (44-20) 7675 1906
e-mail: vijay.verma@orc.co.uk

1. Introduction

Much discussion of sampling methods, including in textbooks on the subject, tends to be confined to the design of population-based surveys, i.e. surveys in which households (or sometimes individual persons) form the ultimate units of selection, collection and analysis. The theory and practice of large-scale population-based sample surveys is reasonably well established and understood.

In this paper I will discuss some special considerations which arise in the design of samples for what may be termed *economic surveys*, as distinct from population-based surveys. By this I mean surveys concerned with *the study of characteristics of economic units*, such as agricultural holdings, household enterprises, own-account businesses, or other establishments in different sectors of the economy. My focus will be exclusively on sample surveys of economic units which, like households, are *small-scale*, *numerous* and *widely dispersed* in the population. Units which are medium-to-large in size, few in number or are not widely dispersed may require different approaches, often based on list frames.

The type of sample designs used in 'typical' household surveys provide the point of departure in this discussion of sampling of other small-scale units. Indeed, there may often be a one-to-one correspondence between such economic units and households, and households rather than the economic units themselves may directly serve as the ultimate sampling units. Nevertheless, despite much common ground with sampling for population-based household surveys, sampling small-scale economic units involves a number of different and additional considerations.

It is useful to begin by noting some similarities between the two situations, and then move on to identifying and developing special features of sampling for surveys of small-scale economic units. At the occasion of this Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR), our special interest here is surveys of agricultural household or holdings, covering diverse types of units, products, activities or sectors of agriculture.

2. Similarities with household survey design

National or other large-scale household surveys are typically based on multi-stage sampling designs. Firstly, a sample of area units is selected in one or more stages, and at the last stage a sample of ultimate units (dwelling, households, persons etc) is selected within each sample area. Increasingly – including, and especially in, developing countries – a more or less standard two-stage design is becoming common. In this design the first stage consists of the selection of area units with probability proportional to some measure of size, (M_k), such as the estimated number of households or persons in area k from some past source providing such information for all areas in the sampling frame. At the second stage, ultimate units are selected within each sample area with probability inversely proportional to size. The overall probability of selection of a unit in area k is

$$f_k = \left(\frac{a.M_k}{M} \right) . \left(\frac{b}{N_k} \right) = f . \left(\frac{M_k}{N_k} \right), \qquad (1)$$

where (a, b, M and f) are constants. Here a is the number of areas selected, when M is the total of M_k values in the population; b is the expected number of ultimate units selected per sample area; hence $a*b=n$ is the expected sample size; and f is a constant defined as

$$f = \frac{a.b}{M} = \frac{n}{M}. \qquad (2)$$

The denominator N_k may be the same as M_k (the measure of size used at the first stage), in which case we get a *self-weighting* sample with $f_k = f = const.$

Alternatively, N_k may be the actual size of the area, in which case we get a 'constant take' design , i.e. with a constant number (b) of ultimate units selected from each sample area irrespective of the size of the area. It is also possible to have N_k as some alternative measure of size, for instance representing a compromise between the above two designs. In any case, M_k and N_k are usually closely related to – and are meant to approximate – the actual size of the area.[1]

It is common in national household surveys to aim at self-weighting or approximately self-weighting designs. This often applies at least within major geographical domains such as urban-rural or regions of the country.

The selection of ultimate units within each sample area requires a listing of these units. Often existing lists have to be updated or new lists prepared for the purpose to capture the current situation. No such lists are required for areas not selected at the first stage. The absence of up-to-date lists of ultimate units for the whole population is a major reason for using area-based multi-stage designs.

Now let us consider a survey of small-scale economic units such as agricultural holdings or other types of household enterprises in similar circumstances. Just like households, such units tend to be numerous and dispersed in the population. Indeed, households themselves may form the ultimate sampling units in such surveys, the economic units of interest coming into the sample through their association with households. Similar to the situation with household surveys, typically no up-to-date lists of small-scale economic units are available for the entire population. This requires resorting to an *area-based multi-stage design*, such as the one implied by (1) above.

---

[1] Throughout, 'size' refers to the number of ultimate units in the area, not to its physical size.

The main difference is that economic surveys generally require major departures from self-weighting designs. We explore and discuss this fundamental difference in the remainder of this paper.

3. Special features of economic surveys

Despite similarities noted above, there are certain major differences in the design requirements of population-based household surveys and surveys of small-scale (often household-based) economic units. These arise from differences in the type and distribution of the units and in the reporting requirements.

3.1 Heterogeneity

Household surveys are generally designed to cover the entire population uniformly. Different subgroups (such as households by size and type, age and sex groups in the population, social classes etc) are often important analysis and reporting categories, but (except possibly for geographical classes) are rarely distinct design domains. By contrast, economic units are characterised by their heterogeneity and by much more uneven spatial distribution. The population comprises of multiple 'sectors', often with great differences in the number, distribution, size and other characteristics of the units in different sectors – representing different types of economic activities to be captured in the survey, possibly using different questionnaires and even data collection methodologies. Separate and detailed reporting by sector tends to be a much more fundamental requirement, than it is in the case for different population subgroups in household surveys. The economic sectors can, and often do, differ greatly in size (number of units in the population) and in sample size (precision) requirements, and hence in the required sampling rates. Therefore it is necessary to treat them not only as separate analysis and reporting categories, but also as *distinct design domains*.

3.2 Uneven distribution

These aspects are accentuated by uneven geographical distribution of economic units of different types. Normally, different sectors to be covered in the same survey are distributed very differently across the population: varying from (i) some sectors concentrated in a few areas, to (ii) some widely dispersed throughout, but with (iii) many 'mixed' sectors concentrated or dispersed to varying degrees. These patterns of geographical distribution have to be captured in the sampling design. True, population subgroups of interest in household surveys can also differ in their distribution (as in the typology of 'geographical, 'cross' and 'mixed' subclasses defined by Leslie Kish in the analysis of design effects), but normally type (ii) rather than (iii) predominate there. In contrast to this, often situation (iii) predominates in economic surveys; and furthermore, as noted above, such 'mixed' sectors need to be treated as distinct design domains.

3.3 Sampling versus survey units

There are a number of other factors which make the design of economic surveys more complex than that of household surveys. Complexity arises from the possibility that the ultimate units used in sample selection may not be of the same type as the units involved in data collection and analysis. The two types of units may lack one-to-one correspondence. For instance, the ultimate sampling units may be (often are) households, each of which may represent none, one or more than one type of economic activity of interest. For instance the same household may undertake different types of agricultural activities. Hence, seen in terms of the ultimate sampling units (households), different sectors (substantive domains) are not disjoint but overlapping. This gives rise to two possible design strategies: (1) An integrated design, based on a common sample of households, in which all sectors of activity in which a selected unit is engaged in would be covered simultaneously. (2) Separate sectoral designs, in which the sector populations (in terms of the sampled units, households) and hence the samples may overlap. In each sectoral survey, activity of the selected households pertaining only to the sector concerned would be enumerated.

Separate surveys are generally more costly and difficult to implement. The overlap between the sectoral samples may be removed by characterising the sampling units (households) in terms of their *predominating* sector. This helps to make the sampling process more manageable. However, this precludes separate sectoral surveys: in so far as the sample for any particular sector is restricted only to the households in which that sector predominates over all other sectors, the coverage of the sector remains incomplete.

3.4 Sampling different types of units

In practice it is often costly, difficult and error-prone to identify and separate out the ultimate survey units into different sectors and apply different sampling procedures or rates by sector. Hence it is desirable, as far as possible, to *absorb any differences in the sampling requirements by sector at preceding area stage(s) of sampling*, so as to avoid having to treat different types of units differently at the ultimate stage of sampling. Of course, the cost of such (very desirable) operational simplification is the increased complexity of the design this may involve.

4. Description of the basic design

Let us now consider some basic features of an area-based multi-stage sampling design for a survey of small-scale economic units. The population of units comprises a number of 'sectors', such as different types of holdings, agricultural activities or products. Sample size requirements have been specified for each sector. The available sampling frame consists of area units (which form the primary sampling units (PSU's) in a two-stage design, or the 'ultimate area units' (UAU's) in a design with multiple area stages). Information is available on the expected number of economic units ($N_{i,k}$) in each area k by sector i, and hence on their total ($N_{.k}$) for the area.[2]

In essence, the overall selection equation (following equation [1] above) is of the form:

[2] Note that this information requirement by sector is more elaborate than a single measure of population size normally required for probability proportional to size (PPS) sampling in a household survey. It is important that, especially in the context of developing countries with limited administrative sources, potential sources such as population, agricultural and economic censuses are designed to yield such information, required for efficient design of surveys of small-scale economic units.

$$f_k = f.\left(\frac{M_{.k}}{N_{.k}}\right), \tag{3}$$

where $M_{.k}$ is some measure of size assigned to the area in its PPS selection, and the selection of ultimate units within the area is with probability inversely proportional to the size measure N_k in the frame, assumed to estimate the actual size of the area.
The design weight to be applied at the estimation stage is inversely proportional to the overall selection probability:

$$w_k = \frac{1}{f_k} \ \alpha \ \left(\frac{N_{.k}}{M_{.k}}\right). \tag{4}$$

The *expected* number of units contributed to the sample by area k is

$$f_k.N_{.k} = f.M_{.k}. \tag{5}$$

Note that the reference here is to the expected value of the contribution of the area to the sample. The actual number of units contributed by any area will be zero if the area is not selected at the preceding stage(s), and generally much larger if the area has been selected.
Summing over all areas k in the population in (5) gives the total expected sample size

$$n = f^*M. \tag{5a}$$

The expected number of units of a particular sector i contributed to the sample by area k is

$$f_k.N_{i,k} = f.M_{.k}.\left(\frac{N_{i,k}}{N_{.k}}\right) = f.M_{.k}.P_{i,k}, \text{ say.} \tag{6}$$

The basic design problem is to determine the 'modified' size measures $M_{.k}$ such that the sample size requirements

$$n_{i.} = f.\Sigma_k\left(M_{.k}.P_{i,k}\right), \tag{7}$$

or in terms of relative quantities more convenient for numerical work

$$\frac{n_{i.}}{n} = \Sigma_k\left(\frac{M_{.k}}{M}\right).P_{i,k}, \tag{7a}$$

are satisfied for all sectors i simultaneously in the most efficient way. The criterion of 'efficiency' also need to be defined (see next section).
In (7) the sum is over all areas in the *population* (and not merely the sample). Note also that the above formulation assumes that at the ultimate sampling stage, units within a sample area are selected at a uniform rate, inversely proportional to $N_{.k}$, irrespective of

the sector. *This is a very desirable feature of the design in practice*. As noted earlier, it is often costly, difficult and error-prone to identify and separate out the ultimate survey units into different sectors and apply different sampling procedures or rates by sector. The preceding sampling stages are assumed to absorb any difference in the required sampling rates by sector through incorporating those in the definition of the size measures $M_{.k}$, yet to be defined.

The most convenient (but also the most unlikely) situation in the application of (7) is when units of different types (sectors) are geographically completely segregated, i.e. when each area contains units belonging to only one particular sector. Indicating by k(i) areas in the set containing units of sector i only (and of no other sector), it can be seen that (7) is satisfied by a simple relationship of the form

$$M_{.k(i)} = g_i . N_{.k(i)}, \tag{8}$$

giving from (7)

$$n_{i.} = f . g_i . \Sigma_{k(i)} N_{.k(i)} = f . g_i . N_{i.}$$

or

$$g_i = \frac{(n_{i.}/N_{i.})}{f} = \left(\frac{f_i}{f}\right), \tag{9}$$

i.e., inflation of all size measures $N_{k(i)}$ in proportion to the required overall sampling rate for the sector.

It is useful to note that once the size measures are so inflated, the required differences in the sectoral sampling rates are automatically ensured and it is no longer necessary to apply the sample selection operation separately sector by sector. The use of *adjusting size measures to simplify the selection operation* is a useful and convenient device, applicable widely in sampling practice.

In reality the situation is more complex because areas generally contain a mixture of units of different types (sectors), and a simple equation like (8) cannot be applied. Clearly, we should inflate the size measure (and hence inflate the selection probabilities) for areas with proportionately more units from sectors which need to be over-sampled, and vice versa. These considerations need to be quantified more precisely. We know of no exact or theoretical solutions to (7), and have to seek empirical (numerical) solutions determining M_k for the PPS selection of areas, solutions which involve trial and error and defy strict optimisation.

5. Evaluation criterion: the effect of weights on sampling precision

5.1 The effect of 'random' weights

The design effect, which measures the efficiency of a sample design compared to a simple random sample of the same size, can be decomposed under certain assumptions into two factors:

- the effect of sample weights, and

° the effect of other aspects of the sample design, such as clustering and stratification.

We are concerned here with the first component – the effect of sample weights on precision. This effect is generally to inflate variances and reduce the overall efficiency of the design. This arises from the difference between the actual size measures N_k, and the modified size measures M_k used in the selection of area units with the objective of meeting the sample size requirements $n_{i.}$ by sector. The increase in variance depends on the variability in the selection probabilities or the resulting design weights (equation (4)). We use the following equations to compare different choices of the modified size measures in terms of their effect on efficiency of the resulting sample in an empirical search for the best, or at least a 'good', solution.
Such decomposition is possible when, as in the present case, weights are 'external' or 'arbitrary', i.e. essentially uncorrelated with population variances. It has been established theoretically and empirically that the effect of such weighting tends to persist uniformly across estimates for diverse variables and population subclasses, including estimates of differentials and trends, and is well approximated by the following expression:

$$D^2 = \left(1 + cv^2\left(w_j\right)\right), \tag{10}$$

where $cv(w_j)$ is the coefficient of variation of the weights of the ultimate units in the sample. The expression approximates the factor by which sampling variances are inflated, i.e. *the effective sample size is reduced*.
From weights w_j for individual units j in the sample of size n, the above can be written as follows, with Σ representing the sum over units in the sample:

$$D^2 = \left(\frac{\Sigma w_j^2}{n}\right) \Big/ \left(\frac{\Sigma w_j}{n}\right)^2 , \tag{11}$$

or, for sets of n_k units with the same uniform weight w_k the above becomes:

$$D^2 = \left(\frac{\Sigma n_k . w_k^2}{\Sigma n_k}\right) \Big/ \left(\frac{\Sigma n_k . w_k}{\Sigma n_k}\right)^2 . \tag{12}$$

5.2 Computation of D^2 from the frame

It is useful to write the above equations in terms of weights of units in the *population*, so that different design strategies can be evaluated *without actually having to draw different samples*:

$$D^2 = \left(\frac{\Sigma\left(1/w_j\right)}{N}\right) . \left(\frac{\Sigma\left(w_j\right)}{N}\right) , \tag{13}$$

or, for sets of N_k units with the same uniform weight w_k the above becomes:

$$D_i^2 = \left(\frac{\Sigma(N_{i,k}/w_k)}{\Sigma N_{i,k}} \right) . \left(\frac{\Sigma(N_{i,k}.w_k)}{\Sigma N_{i,k}} \right), \qquad w_k = \frac{N_{.k}}{M_{.k}} \tag{14}$$

The above equation applies to the total population, as well as separately to each sector *i*. ideally, subsampling *within* any area k is identical for all sectors (i), implying uniform weights w_k for all types of units in the area. These weights are proportional to (N_k/M_k) for the area concerned.

The loss due to weighting also has an effect on the effective allocation by sector actually achieved. The effective sample size $n'_{i.}$ in the presence of arbitrary weights is smaller by the factor ($1/D^2_i$), compared to the actual sample size $n_{i.}$. The average of D^2_i values over the sectors

$$\frac{\Sigma_i D_i^2}{I_i} \tag{15}$$

may be taken as an *overall indicator* of the inflation in variance due to weighting in the sectoral designs. The objective is to minimise this indicator by appropriately choosing the size measures M_k satisfying the required sample allocation (7).

6. Constructing "strata of concentration" (StrCon)

To adjust the measures of size M_k and hence the overall sampling rates to achieve the required sample size by sector, it is useful to begin by classifying areas into groups according to which particular sector *predominates* in the area. The basic idea is as follows.

For each sector, the corresponding 'stratum of concentration' is defined to consist of a set of areas in which that sector 'predominates' in the sense as defined below. One such stratum corresponds to each sector. The objective of constructing such strata is to separate out areas according to their composition in terms of sector. (To distinguish these from 'ordinary' strata used for sample selection, I will henceforth refer to them as "StrCon".)

6.1 Notation

- i subscript identifying a sector, i=1 to I.
- k subscript identifying an area
- $N_{i,k}$ number of units of sector i in area k
- $N_{.k}$ total number of units in area k (all sector combined)
- $N_{i.}$ total number of units of sector i (all areas in the population)
- A_i number of areas containing at least 1 units of sector i ($N_{i,k}$>0)
- B_i average number of units of sector i per area (counting only areas containing at least 1 such unit) = ($N_{i.}$ / A_i)
- $P_{i,k}$ 'index of relative concentration' of sector i in area k = $N_{i,k}$ / B_i
- j index (1-3) identifying the stratum of concentration (StrCon), i.e. the sector i which has the highest $P_{i,k}$ value for the area k

Note that the 'index of relative concentration' $P_{i,k}$ has been defined in relative terms: the number of units of a particular sector i in area k in relation to the average number of such units per area. Defining this simply in terms of the actual number of units will result in automatic over-domination of the largest sectors. However, for a sector not large in overall size but confined to only a small proportion of the areas, the average per area (including zeros) would tend to be small, resulting in *its* over-domination. Hence it appears appropriate to exclude zeros (areas with $N_{i,k}=0$) in computing the average B_i.

6.2 Data by StrCon and sector (aggregated over areas)

Once the StrCon has been defined for each area, the basic information aggregated over areas on the numbers of units classified by sector and StrCon can be represented in a square table of the following form (Table 1).

There is a one-to-one correspondence between the sectors and the StrCon. Subscript i (rows 1 to I) refers to sector, and j (columns 1 to I) to StrCon. Generally, any sector is distributed over various StrCon and any StrCon contains units from various sectors. However, the DIAGONAL elements (i=j) predominate to the extent sectors are geographically segregated, and each tends to be concentrated within its 'own' StrCon.

Table 1: *Classification of units by sector and the 'strata of concentration'*

StrCon	1	…	j	…	I	total
Sector						
1	i=j					
…		i=j				
i			Nij			Ni., ni.,fi
				i=j		
I					i=j	
total			N.j			N.., n..,f

The following notation is used in the table.

Nij refers to the number of units of sector i in all areas in StrCon j. Summing along rows,

Ni. is the total number of units in sector i, $\Sigma_j N_{ij}$. Summing along columns,

N.j is the number of units in StrCon j, $\Sigma_i N_{ij}$.

ni. is the target sample size for sector i,

fi = (ni. / Ni.) is the implied overall sampling rate for sector i.

$N_{k(j)}$ where necessary, k(j) will be used to refer to a particular area k in StrCon j. Hence for instance, summed over areas in the frame for StrCon (j): $N_{ij}= \Sigma_{k(j)} N_{i,k(j)}$, $N_{.j}= \Sigma_{k(j)} N_{.k(j)}$

7. Using StrCon for determining the sampling rates

7.1 A basic model

The use of StrCon as defined above is the fundamental aspect of the strategy for assigning measures of size to areas for the purpose of achieving the required sample allocation by sector.

We will first consider a simple model using a uniform sampling rate within each StrCon, but varied across StrCon with the objective of obtaining the required allocation by sector. This basic model will be refined subsequently.
The model implies the assignment of area measures of size in the form

$$M_{.k(j)} = g_j . N_{.k(j)} , \tag{16}$$

where the actual size N of area k in StrCon j is inflated by a factor g_j constant for the StrCon. This is the factor by which the ultimate selection probabilities of units in the area are inflated. Note that in (16) the scale of M and hence of g_j is arbitrary. For numerical convenience, it can be chosen such that for (16) summed over the whole frame, M=N.
Since the above inflation in this basic model applies equally to units of all sectors in an area, it can be written in terms of cells of Table 1 as

$$M_{ij} = g_j.N_{ij} . \tag{17}$$

Equation (16) looks formally the same as (8); but the latter is a particular and very simple case of (16), for the situation when Table 1 has non-zero elements only in the diagonal (representing complete geographical separation of different sectors).
Suppose that the objective is to obtain sample size $n_{i.}$ for sector i, i.e. to apply the average sampling rate

$$f_i = (n_{i.} / N_{i.}). \tag{18}$$

This rate cannot, of course, be applied directly to rows of Table 1 (i.e. to numbers of units by sector), but must be applied to (whole) areas, i.e. along columns of the table. The basic requirement is to solve the following set of linear equations to determine the StrCon sampling rates say $(f.g_j)$ - or its exact equivalent, the required adjustment to measures of size (10) - given the distribution N_{ij} and the target sample sizes $n_{i.}$ by sector i:

$$f.\Sigma_j\left(M_{ij}\right) = f.\Sigma_j\left(g_j.N_{ij}\right) = f_i.N_{i.} = n_{i.} \text{, for j,i = 1 to I.} \tag{19}$$

or in terms of relative quantities more convenient for numerical work

$$\frac{n_{i.}}{n} = \Sigma_j g_j.\left(N_{ij}/N\right). \tag{19a}$$

In principle, the above can be solved by inverting an (IxI) matrix. However, an iterative solution can be simpler and more convenient. To the extent diagonal elements predominate in Table 1, a good starting point is to assume $g_j=(f_i/f)$ for j=i, and then iteratively adjust the g_j values to satisfy (19) within a certain margin.

7.2 More flexible models: an empirical approach

Depending on the numbers and distribution of units of different types and on the extent to which the required sampling rates by sector differ, a basic model like (16) may be too inflexible, and may result in large variations in design weights and hence large losses in efficiency of the design. It may even prove impossible to satisfy the sample allocation requirements.

Lacking a general theoretical solution, we have tried more flexible empirical approaches to defining the modified size measures to meet the required sample allocation and to achieve this more efficiently.

Basically, the approach has involved supplementing (16) by further modifying the size measures in a more *targeted* fashion, as follows.

The original size measure broken down by sector is

$$N_{.k(j)} = \Sigma_i N_{i,k(j)} = \left(N_{.k(j)} - N_{j,k(j)}\right) + \left(N_{j,k(j)}\right), \qquad (20)$$

and the modified measure is defined in the form

$$M_{.k(j)} = g_j.\left[\left(N_{.k(j)} - N_{j,k(j)}\right) + h_{j,k(j)}.\left(N_{j,k(j)}\right)\right], \qquad (21)$$

The first term on the right in (20) is the original size measure in terms of all units other than those belonging to the sector (i=j) corresponding to the area's StrCon (j). The second term is the original size measure for units belonging to that sector. It is the latter which is modified by some factor h in (21) to facilitate meeting the sample allocation constraints. The variation in the overall sampling rates is determined by the ratio (21)/(20):

$$\frac{M_{.k(j)}}{N_{.k(j)}} = g_j.\left[(1-x) + h_j(x).x\right], \qquad (22)$$

where 'x' stands, for a given area k, for the proportion of units belonging to the sector (i=j) which corresponds to StrCon (j) of the area:

$$\frac{N_{j,k(j)}}{N_{.k(j)}} = P_{j,k(j)} = x, \quad say. \qquad (23)$$

It is assumed in (22) that an appropriate form for h_j is to take it as a function of x; subscript j indicates that form may differ by StrCon.

The above can be aggregated over all areas k(j) within each StrCon to construct cells of Table 1:

$$N_{ij} = \Sigma_{k(j)} N_{i,k(j)},$$

$$M_{ij} = \Sigma_{k(j)}\left(\frac{M_{.k(j)}}{N_{.k(j)}}\right)N_{i,k(j)} = g_j.\Sigma_{k(j)}\left((1-x) + h_j(x).x\right)N_{i,k(j)} = g_j.N'_{ij}, \text{ say.} \qquad (24)$$

The sample allocation constraint is

$$f.\Sigma_j\left(M_{ij}\right)=f.\Sigma_j\left(g_j.N'_{ij}\right)=n_{i.}\text{, for j,i = 1 to I,} \tag{25}$$

or in terms of relative quantities more convenient for numerical work

$$\frac{n_{i.}}{n}=\Sigma_j g_j.\left(N'_{ij}/N\right). \tag{25a}$$

Our empirical approach has involved the following steps:

- choosing a form for *h*, and using it to determine N'_{ij} defined in (24)
- using those in (25a) to iteratively determine the g_j values by StrCon which meet the sample allocation requirements $n_{.i}$ by sector, and hence the M_{ij} values from (24)
- computing the implied losses in efficiency due to weighting from (14), and their average over sectors, (15)
- comparing this average against (many) other choices of the function h, and choosing the most efficient solution from among those computed.

The objective is to identify and choose the 'best' model, at least among those empirically evaluated. Comparing large numbers of numerical trials points at least to *the direction we should be moving in* in the choice of the design parameters.

7.3 Illustrations

Here are examples of some of the forms which we have investigated and compared:

Basic: h=1, simply reducing (22) to its basic form (16):

$$\frac{M_{.k(j)}}{N_{.k(j)}}=g_j\,.$$

Constant: The size measure of units of the sector corresponding to the area's StrCon adjusted by a constant factor:

$$\frac{M_{.k(j)}}{N_{.k(j)}}=g_j.\left[(1-x)+\left(1+c_j\right).x\right].$$

Linear: The above size measures adjusted in linear proportion to x, the proportion of units in the area belonging to the sector corresponding to the area's StrCon:

$$\frac{M_{.k(j)}}{N_{.k(j)}}=g_j.\left[(1-x)+\left(1+c_j.x\right).x\right],$$

S-shaped: The above with more elaborate variation with x; for instance:

$$\frac{M_{.k(j)}}{N_{.k(j)}} = g_j.\left[(1-x)+\left(1+c_j.x^2.(3-2x)\right).x\right],$$

and so on. The last mentioned form appeared a good one at least for one survey tried. The models above all assume $h_j(x)$ to be of the form

$$h_j(x)=c_j \,.\, h(x),$$

where function h(x) is independent of j, i.e. is taken to be the same in all sectors in a given trial of the procedure. As for parameter c_j in the above, a simple choice we have tried is to take it as a function of the overall sampling rate required in the sector (i=j) corresponding to StrCon (j), such as:

$$c_j = \left[\left(\frac{f_i}{f}\right)^{\alpha} - 1\right], \quad \alpha \geq 0, \quad i = j \,.$$

8. Concluding remarks

The specific examples given above must be taken as merely illustrative of the type of solutions we have looked into to the basic design problem in surveys of heterogeneous and unevenly distributed small-scale economic units. The solution depends on the nature of the population at hand, and we have used the approach sketched above in designing a number of samples, covering diverse types of surveys in different situations (countries). These have included surveys of agricultural and non-agricultural small-scale units, and even in a survey of schools where the objective was to control the sample allocation by ethnic group but through a sample of schools on the basis of information on the schools' ethnic composition.

In this last example, it was not ethically permissible to identify and sample differentially students of different ethnic groups: all variations in the required overall sampling rates by ethnicity had to be achieved by appropriately adjusting the school selection probabilities (i.e. measures of size) according to pre-available information on the schools' ethnic composition.

In conclusion, one important and useful aspect of the above approach should be brought out. Meeting sample allocation requirements is a fundamental aspect of the design for surveys of the kind under discussion. Nevertheless, to the extent possible, the issue of *sample allocation* – involving the choice of the modification of measures of size (M_k/N_k) for instance – should be isolated from the structure and process of actual *sample selection*. Once assigned, the units always 'carry with themselves' their relative selection probabilities (M_k/N_k) irrespective of details of the selection process, and the required allocation is automatically ensured, at least in the statistically expected sense.

The advantage of such separation of allocation and selection aspects is that the structure and process of selection (stratification, multiple sampling stages, subsampling etc) can be determined flexibly, purely by considerations of sampling efficiency, and need not be constrained by the requirements of allocation.

When the two aspects overlap in practice – for instance when a whole design domain is to be over-sampled – that is a coincidental rather than an essential aspect of the design.

7 June 2001

Invited Paper Session

NEW TECHNOLOGIES AND STRATEGIES FOR AGRICULTURAL DATA COLLECTION AND FOR USE IN ENVIRONMENT MONITORING

Chair: D. Heath

Macro-Level Interactions Between Agriculture and Environment : Integration of Statistical and Geographical Information Systems

Jean-Louis Weber
Ifen, French Environment Institute, 61, Bd Alexandre Martin, F-45058 Orléans, France
e-mail : jean-louis.weber@ifen.fr

Abstract: It is now possible to complement the statistical assessment of the agriculture production system and the statistical reports on the state of environment by the overall description of the way they interact at various scales. When it was formerly just possible to reflect on isolated cases and/or on hypothetical correlation between aggregated statistics, it is now possible to produce and analyse statistics based on a direct approach of the local interactions. For the purpose of producing indicators, this statistic will be aggregated at the meso and macro levels, according to the landscape patterns to which it is relevant. The integration of statistical and geographical information covers sampling to modelling, compilation of indicators and environmental accounting of the land. It is of relevance for new policies related to the development of the rural landscape in terms of economic performance as well as for the maintenance of its ecological functions.

Keywords: Environment, Agriculture, Sampling, Modelling, Indicators, Spatial Analysis, Geographic Information System, Environmental Land Accounts.

1. Introduction

Since the origin, the natural environment of agriculture is an obvious condition of the farmer's activity and way of life. However, due to the fast development of agriculture in the past century, geared by the necessity to feed the increasing urban market, the agro-system has been regarded as a production system, the emphasis being put on the production of products. Stocks, crops, yields, prices turned to be the key indicators of agriculture efficiency. Environmental conditions at this stage appeared mainly as limiting factors.

More recently, regional or local environmental problems, as well as of the understanding of the concept of sustainable development, lead to a new questioning of the interactions between agriculture and the environment. This new character is not related so much to the understanding of the issue as such. Agronomists as well as environmental scientists have always analysed the impacts of agriculture practices on water resources, the quality of the soil, the botanical diversity etc. But their primary purpose was, in general, the identification of best practices to be implemented at the farm level or in local patterns, rather than an overall assessment of the interaction between agriculture and the environment.

At the same time, programmes of collection of environmental statistics started at the OECD, the UN and Eurostat. But theses statistics were, in most cases, too much aggregated to feed the implicit causal model for which they have been collected (the so-called Pressure-State-Response model). At this stage, the spatial distribution of the phenomenon was, to a large extent, ignored, which created serious limitations for analysing the statistical distribution of the interactions. Consequently, it was difficult to assess, at the macro-level, the environmental impacts of agriculture development and to supply the policy makers with the information that they could have required in this domain.

Things started to change when, on the one hand, the pressure of human activities on the environment and the natural resource reached a threshold and, at the same time, when observation and data handling tools offered the possibility of producing information for new policies. The threshold can be set with the opening of the markets of the products, with its consequences in terms of intensification of the productions as well as of the development of transport and increasing specialisation of regions or areas according to their comparative capacities. The technological change is earmarked by earth observation satellites, geographical information systems, as well as the development of telecommunications which made feasible the implementation of real time monitoring systems.

It is therefore possible to complement the statistical assessment of the agriculture production system and the environmental statistical reports with the overall description of the way interactions between agriculture and the environment occur at various scales. Where it was formerly just possible to reflect on isolated cases and/or on a hypothetical correlation between aggregated statistics, it is now possible to produce and analyse a statistic based on a direct approach of the local interactions. This statistic will be aggregated at the meso and macro levels, according to the landscape patterns to which it is relevant.

The strong demand by the policy makers for an integration of the environment in the sectoral policies at the European Union level, initiated by the so-called Cardiff Initiative ("Putting Sustainable Development in the Fast track", 1997) and continued with the orientations of the Goteborg Summit (June, 2001) has a counterpart in terms of the information requested. It is commonly presented according to the "two corridors model" leading to an integrated monitoring and reporting system which is foreseen. The system is based on two packages of indicators. The first one relates to the implementation, follow up and assessment of the 6^{th} Environmental Action Plan and of its specific developments (water, nature...). The second package of indicators deals with sectoral strategies and action programmes. Priority sectors are agriculture, transport, energy. This European approach is in line with other international developments, for example with the OECD project on agri-environmental indicators.

Therefore, developments involving a multidisciplinary co-operation are many. They lead to a double cross fertilisation :

- between environmental and agricultural perspectives;
- between statistical and geographical approaches.

They can be classified under four items : sampling, modelling, development of indicators, accounting. The presentation below is based on examples. Some will better illuminate the agri-environmental integration, others the combined use of statistics and geographical tools.

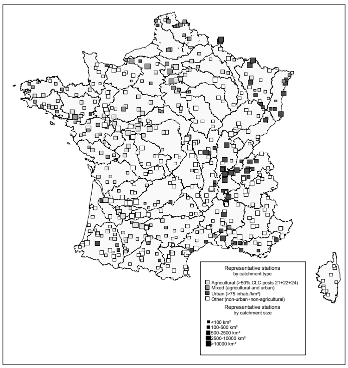
2. Geographical stratification in sampling procedures

The purpose of this section is not to address issues related to aerial frame sampling in general. Geography is here defined in a restricted sense which is the description of the distribution and of the nature of the phenomenon. Therefore, surveys which aim at producing statistics from samples distributed over the territory at random or systematically (regular grid), are only considered in their relation to geographic information.

2.1. Eurowaternet (EEA) : stratification of the monitoring system of the quality of the rivers for producing representative samples

The implementation of the EUROWATERNET river quality network of the European Environmental Agency requires progress in terms of comparability of the data collected from the present monitoring networks. Beside problems of inter-calibration of the measurements, a crucial issue is the statistical representativity of the data sets at the national and river basin levels. This implementation is carried out applying the methods of designing elaborated by the European Topic Centre on Waters of the EEA.

Figure 2 *: The French baseline Eurowaternet monitoring network of the rivers (representative at the national level, for N & P) (Leonard & Crouzet, 1999)*



Before setting up guidelines, a test study has been worked out in France (Leonard and Crouzet, 1999). The proposed network, representative of the presence of nutrients in rivers at the national level, is made of 552 sampling stations selected according to selection criteria calculated thanks to indicators derived from CORINE Land Cover and other sources.

Selection criteria to characterise rivers in order to ensure the best representativity with respect to river types, basin size (surface, run off), pressure by human activities. As long as the run off which measures the size of the basin is integrating the flow coming from upstream, the human pressure has been taken into account for each elementary basin for an amount cumulating the corresponding (upstream) factors.

In the case of Nitrogen and Phosphorus, the sources of pressure to be considered have been limited to agriculture and population. CORINE Land Cover, through the HYDROSOL data model, has been used to calculate the cumulated indicators of agricultural pressure along the river systems delimited by the c.a. 6200 elementary catchments of metropolitan France (the “hydrographic zones”).

Urban pressure has been assessed in the same way, from data on population density. A map of combined pressure is produced, with four classes : “urban”, “agricultural”, “mixed” (Urban AND agricultural), “other” (none of these pressures being a substantial threat).

The selection of the repesentative sample takes into account the national coverage, the size of the elementary catchment and the proportion of pressure class. The selected stations allow calculation of statistics representing, in the test study, the annual mean of nitrate (NO_3), ammonium (NH_4) and soluble reactive phosphorus (PO_4).

2.2. The use of Corine Land Cover to improve area frame surveys; example of MARS (JRC)

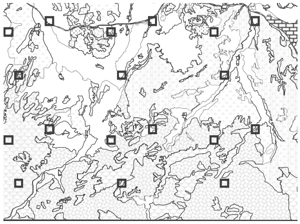
Corine Land Cover is the inventory (map and database) realised by the EEA in co-operation with the European Commission and the member countries. 30 countries are now covered at the scale of 100 000, with a standard classification of 44 items structured over 3 levels. The methodology is based photo-interpretation of satellite images with the support of ancillary information (aerial photos, image processing for identification of particular items, current maps...). The size of the smallest land cover unit is 25 hectares. The first CLC relates more or less (according to countries) to the year 1990. The updating process is now ongoing under the name of CLC2000, based on satellite images taken between 1999 and 2001.

CLC identifies agriculture land cover units. However, its purpose is more a consistent description of landscape patterns than a precise assessment of the surface of each land cover type. More, because of its analytical purpose and of the methodology used (photo-interpretation with a minimum unit of 25 hectares), the CLC land cover units (which are zones and not parcels) are not 100 % homogenous. More over, the CLC nomenclature recognises explicitly the existence of mixed classes such as " Discontinuous Urban Fabric", " Annual Crops Associated With Permanent Crops", "Complex Cultivation Patterns", "Land Principally Occupied By Agriculture, With Significant Areas Of Natural Vegetation", "Agro-Forestry Areas" or " Sparsely Vegetated Areas". From an ecosystemic point of view, these types are relevant and provide very interesting indicators (*see 6.2, below*).

Of course, in the perspective of an accurate assessment of the area of cultivated surfaces, as required for the production of annual statistics on crops, CLC provides a coarse estimation. In addition, CLC is to be updated every 10 years only. CLC cannot be used for a direct estimate of the agriculture surfaces. However, CLC can be efficiently used for improving area frame surveys.

Additional studies have been carried out in the context of the MARS project of JRC "*to overcome the problems of using CORINE Land Cover as a basis for stratification of an area frame of square segments and assess the potential relative efficiency*" (Gallego, 1999).

Figure 3 : *Sample of segments overlaid on CORINE Land Cover classes (Gallego et alii, 2000)*



Results can be read in " The use of CORINE Land Cover to improve area frame survey estimates in Spain" (Gallego et al., 2000).

The authors conclude that *"CORINE Land Cover is a valuable tool to improve the efficiency of land cover area estimates from an area frame ground survey. However it should not be used, as in general any land cover map, for direct area estimation as the sum of the area of polygons of a certain land cover class. Some users may be tempted by the easy use of a land cover map in a GIS framework and take the total area of CORINE Land Cover polygons labelled as "Arable land" or "Vineyards" as an estimate of the*

area of arable land or vineyards. A more consisting approach is combining the exhaustive information, with a relatively coarse scale, with more accurate information coming from a ground survey on a sample of area elements, often named "segments".
"Using CORINE Land Cover classes, possibly grouped into a few classes, as strata is complicated to manage and does not seem to yield a very good relative efficiency. An easy way to use CORINE Land Cover as a covariable for area estimates of the main annual crops is computing an index that approximately corresponds to the area of arable land in each unit of the sampling frame."(Gallego, 1999)

The stratification efficiency index compares, the number of segments requested from a simple random sample to obtain the same results than with a stratified sample of 1000 segments (with an equivalent accuracy). In the case of Spain, it reads : Although the gain in efficiency is smaller than with ground survey of surfaces, it indicates "*a good potential of Corine land cover as a basis for stratification*." (Gallego, 1999)

Table 1: *Efficiency of a stratification of square segments based on CORINE Land Cover in Spain (Gallego, 1999)*

	Efficiency
Wheat	2.21
Barley	1.51
Cereals	2.18
Sunflower	1.74
Sugar beet	1.74
Fallow	1.18

2.3. Countryside Survey (GB) : sampling and typology of landscape

The Countryside Survey is a comprehensive periodic census of the rural landscape of Great Britain (Countryside survey 2000). Its main focus is the assessment of the many aspects of the biological diversity in relation with the human activities which influence it, such as agriculture or the management of the landscape itself. The information collected through the survey forms part of the environmental reporting system of the Department of Environment of the UK.

Figure 4 : *The Four Major Landscape Types of Great Britain used for Environmental Accounts (after Barr* et al. *1993).*

The two last versions of the Countryside Survey, so-called CS1990 and CS2000, are based on the same general approach, although some methodological changes have taken place in order to improve such or such element of the system. Basically, the CS methodology combines sampling and census, the methodology employed being based on a sample of 1 km x 1 km grid squares stratified by broad environmental land classes within GB (Haines-Young, 1997). Remote sensing is used as a first level of monitoring changes in land cover. Land cover data are combined, in a second step, with climate and geo-morphological data for defining environmental land classes. This constitutes the basis for the stratification of the field survey. An aggregated picture of the land classes aggregated in four types can be seen beside.

Sample-based surveys have the advantage that they can provide very detailed information about land cover and the way it is changing. The drawback of these data is that it is not possible to map elements precisely, but only to predict where certain types of conditions may occur. The field survey element of CS1990 and 2000 can be described

as a non-targeted survey in that (apart from urban areas, which were excluded from the survey) the chance of sampling a given cover type is proportional to its area in GB. It shows how sample-based data can be integrated most effectively with census-type information to give a complete picture of land cover and land cover change, than is possible using any single source or type of data.
However, the Policy Review of CS1990 assessed the extent to which existing sampling strategies provide adequate thematic coverage, in terms of producing information about both common and rare land cover types. In the second case, it indicated that they have been neglected by the sampling approach, and that the survey had almost nothing to say about the rare but valuable habitats or land cover types in GB, whose importance in conservation terms far outweighs their area.

"Clearly with any sample-based survey, the chances of including these more infrequent elements could be increased by increasing sample size. In practice, however, this is strategy is rarely an option because of constraints of cost and time.... An alternative approach is to target particular resources of interest. This was done as a survey of key habitats which followed CS1990, in which sampling was targeted on particular types of habitat, namely calcareous grassland, heath, upland, waterside and coastal habitats in England. The targeting was achieved by applying various types of habitat mask to the population of 1 km squares which make up GB, and which were sampled during CS1990... The study suggested that the greatest threat to these habitats is probably no longer loss of area, but the effects of inappropriate management practices on their quality." (Haines-Young, 1997).

The Countryside survey provides a unique set of information for a systematic description of the interactions between agriculture and the environment. It has been broadly used as a basis for compiling environmental accounts which incorporate economic as well as ecological and social aspects and contribute to the reporting on the state of the environment and the sustainable development in the UK. One example is given with the report "Accounting for nature : assessing habitats in the UK countryside" (DETR, 2000).

3. Integration of statistics and geographic data in modelling

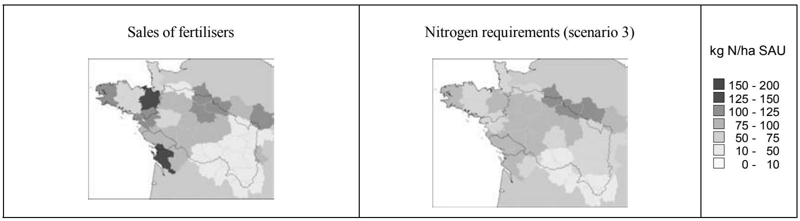
The integration of conventional statistics and geographical data covers a wide range of issues. Only a few examples will be presented in this section, having in mind the storyline : interactions between agriculture and the environment at the macro level. The first example will deal with difficulties in obtaining an appropriate geographical breakdown for data which have to be collected by statistical surveys. The second will illuminate the possible input of spatial analysis in a domain currently studied with demographic statistics. The third example relates to the possibility of extrapolating data and findings of local models by using a geographical stratification. The last one is of a more methodological nature and presents one possibility of geostatistical modelling of landscape features and statistical data.

3.1. Modelling emissions of N and P to rivers

The assessment of emissions of Nitrogen and Phosphorus to rivers requires the combination of data on point sources (the municipalities and industrial plants which discharge waste water in specific points) and non-point sources (the leakage of the surplus of N & P related to the use of fertilisers). When the first type of emission can be monitored, the second can only be calculated through the use of physical modelling. When considering the rivers, the hydrological models are based, for clear reasons, on an analysis by river basins. For good results, these river basins should be of a rather small size, which mean that most (if not all) data must be available at this scale. The definition of the hydrological models as well as the validation has been made at such a scale, that it requires important field surveys, due, inter alias, to the availability of some statistics (fertilisers, manure, fodder...) which are currently collected on a sample basis or with a reference to administrative units (and not physical limits as is required here). Such an approach is not operational at the European or even at the national level for reasons of cost and or timeliness. Solutions have to be found and developments are taking place in Europe (the EuroHarp project) as well as in countries (France, Germany).

For example, in France, the effective use of fertilisers is known through an agriculture census which takes place every 8 to 10 years. Annual estimates are based on the sales of fertilisers. However, annual statistics of sales of fertilisers are inaccurate for running the hydrological model used by Ifen, as long as they are reported at the place of sale, generally in cities and not in the areas where they are used. The problem can be illuminated by the figure below :

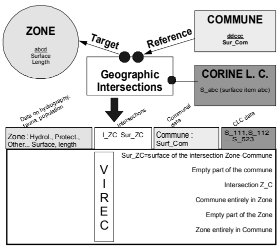
Figure 5 : *Example of gap between statistics of sales of fertilisers by "département" (NUTS3) and their use, estimated from agriculture statistics (Crouzet, 2000)*



To overcome this gap, Ifen used a method of (proportional weighted) reallocation of statistics of sales of fertilisers to small river basins as an estimate of use. The method is based on the surfaces identified as cultivated in Corine land cover. The conceptual model (Hydrosol, Ifen, 1998) is presented on Figure 6.

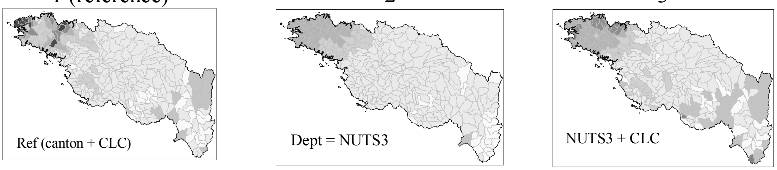
The improvement of the use of the Hydrosol model on the results can be seen by comparing these three maps which show the results in terms of surplus of N. In the first case, the data for fertilisers comes from statistics at the NUTS4 level (in France, "cantons", of which the average dimension is that of municipalities in other European countries). In the

Figure 6 *: The Hydrosol model*



second case, statistics available at the NUTS3 level ("départements") have been directly used as average values. The two pictures are very different. In the 3rd map, input data has been reprocessed with Hydrosol and Corine land cover. The results of map 3 are close to those of map 1.

Figure 7: *Improvement of the results by combining CLC and statistics (Crouzet, 2000)*



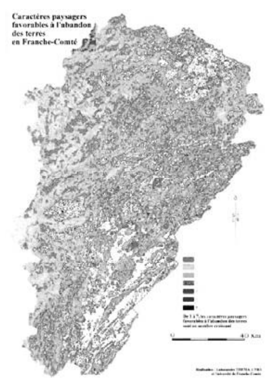
Therefore, the use of CLC allows the use of loose statistics, nevertheless providing valuable results, thus making surplus calculations sustainable. Problems to solve are limited to agronomic know-how definition of scenarios and cross checking procedures. Calculations at European scale hence become affordable on the short term.

3.2. Landscape analysis : Probability of land abandonment by agriculture

Land abandonment by agriculture often happens when the young don't want to continue the operation of their parents' farm. Analysing and anticipating the phenomenon can be done, to some extent, by reference to demographic statistics. Other approaches provide alternative or supplementary sources of information. Land abandonment is the consequence of comparative disadvantage of farms or regions which cannot increase their productivity and/or sell their products. Partly, this disadvantage can be explained by the combination of geographical determinants such as the slope or the remoteness.

A study made by the Thema Laboratory of the French CNRS and the University of Franche-Comté (Joly, Brossard et al., 1997) on the physical characteristics of landscape selected and combined several factors contributing to abandonment of land by agriculture. They defined them in order to be able to work out the analysis from a limited set of geographical data bases: Corine land cover, the roads and the digital elevation model. These factors favourable to land abandonment are the land cover type (mixed agriculture patterns), adjacency to forests, altitude, slope, and distance from main roads. 7 classes of probability are defined, from low to high (light green to dark brown on the map beside).

Figure 8: *Landscape characteristics favourable to land abanonment, Franche-Comté, (after THEMA, 1997)*



The model has been tested successfully with CLC1984 data and the results have been compared with the observation of land cover changes. In this domain, combining economic scenarios with demographic and land cover changes forecasts opens the way to prospects established at the different scales required by policy makers.

3.3. Lucas and CLC : extrapolation of local parameters and models

Sampling on a regular area frame is an indirect way of surveying crops, starting from an accurate estimation of the cultivated surfaces. The LUCAS survey currently implemented by Eurostat belongs to this type.

The statistics derived from such surveys can be very relevant for agri-environmental indicators, specially when details are important in terms of type of crops or of agriculture practice, and when the assessment of trends matters. This information, collected at the marginal cost, can be very useful (and in many cases, unique) for monitoring annual changes in land cover. In addition, specific questions related more directly to environmental issues (wetlands, recreational use) can be introduced in the specifications and reported by the surveyor in the field. The counterpart is that one has to accept the statistical nature of this information, the possible geographical breakdowns being limited by the relative importance of a specific issue. Also, rare land use or cover types will not be adequately surveyed (see the discussion of the Countryside survey in section 2.3).

Another possibility of using regular grid based sampling is modelling. The field survey of the sampling points permits observation of many parameters, ahead of the initial purpose of monitoring cultivated surfaces. Each sampling point can be associated with various attributes from which the classification of land cover and land use types can be derived. Parameters related to natural attributes of land can be integrated, such as soil or vegetation and help in producing an information of high relevance for analysing agriculture systems as well as natural habitats by combining according to the spatial analysis relevant for specific issues.

Potentially, the variety of attributes which can be associated to the sampling points is high and cover issues such as crops, erosion of soil, linear features, buildings and their use, landscape, natural vegetation as well as real images of the landscape. For example, it is foreseen that each sampling point of Lucas is described by photos (N, E, W and S). These photos will be used subsequently for assessing the visual landscape elements associated to the field parameters. The major limitation to data collection remains the skill of the surveyors in the many domains of interest and the additional time required for registering the observations. In practice, this collection of information will be limited but of a very unique nature in itself, because of its association with agriculture databases.

Another advantage of the sampling surveys is their repetitiveness. Crops need to be monitored at an annual pace and environmental statistics can benefit of the initial investment realised for the purpose of agriculture policies. This is particularly important for the assessment of the changes and the process of which their result. Modelling the interaction between agriculture and the environment on points or, better, on clusters of points is a very promising course of research.

However, a question has to be addressed, related, as in the case of the Countryside survey, to the rate of sampling. Local models are interesting for policy makers at the macro level only if their findings can be extrapolated. If the rate of sampling is not sufficient, some post-stratification has to be looked for, in order to produce some statistical results for regions and/or landscape types. These landscape types will relate to geographical characteristics as well as to vegetation and pressure generated by the use of land. GIS techniques and geostatistical analysis of land cover can help in producing zoning(s) and typology(ies) for some generalisation or extrapolation of the results of the local models.

3.4. Smoothing Corine land cover : CORILIS

The CORINE land cover database is currently used for the identification, delineation and characterisation of specific zones such as, e.g. protected areas or for general observation purposes. Georeferenced data about the type of land cover of every point of a territory is a source of information broadly used in geographical information systems.

Smoothing techniques applied to information on land cover make it possible to synthesise its main characteristics at the level of a region or any territory. In doing so, one get rid of the exact land cover of a given spot for encompassing what exists in the neighbourhood. Subsequently, the broad functions of a territory, the understanding of which is essential for land planning, are illuminated

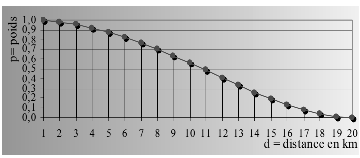
As compared to cartographic generalisation, which consists in eliminating the small features to put the large one in evidence (increasing their original weight by doing so), smoothing methodologies keep track of all the original data and the statistical results are independent of the scale of analysis as well as of the level of aggregation of the calculations. Aggregating smoothed values or smoothing aggregated values leads to the same result.

A methodology for smoothing Corine land cover data has been developed by Ifen under the name of Corilis (Lacaze, 1999) and has been implemented for spatial analysis (Lacaze et al., 2000). An application at the European scale has been worked out for the EEA.

The technique used for Corilis consists in determining for each point of the territory the potential of information included in a determined neighbourhood (radius of circle = *R*).

A statistical function of Gaussian type, so-called «BiWeight» (Chataignon & Grasland, 1998), is used in order to weight this information ($\pi = (1-(d/R)^2)^2$) according to the distance to the determined (*d* kilometres). It can be represented by the following graph, where *p* is a weighting factor.

Figure 9: *Effect of the BiWeight method (Lacaze, 1999)*



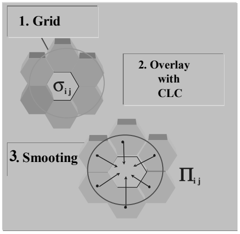
In the first application in France, CLC has been converted in an hexagonal grid (53 497 hexagons with a 2 km base, corresponding to a surface φ of about 10,39 km^2) and the radius for smoothing has been fixed at 20 km (Lacaze, 1999).

The overlay [Corine ∩ Hexagons] gives, for each hexagon j and for each land cover type i, a gross surface (σ i j). The third step is the smoothing calculation [Hexagone × {Hexagons in the 20 km radius}]. In a given hexagon, the smoothed surface (Si j) of each land cover type is determined as the total sum of the corresponding value in the neighbouring cells weighted inversely to their distance :

$$Sij = —\lambda\ (\sigma i\lambda * \pi\lambda) \text{ where } \pi\lambda = 1-(d\lambda/20)^2)^2 \text{ with } d\lambda < 20$$

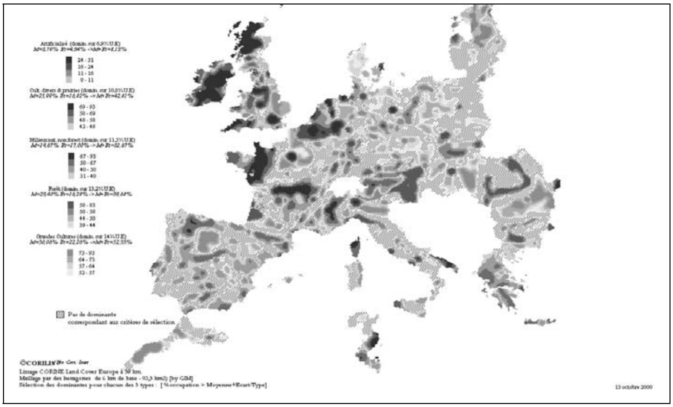
The total smoothed surface being $\Phi_J = —_\lambda (\varphi * \pi_\lambda) = \varphi * —_\lambda \pi_\lambda$, it is possible to calculate the density of each type of land cover in each hexagon ($\Pi_{ij} = —_\lambda (\sigma_{i\lambda} * \pi_\lambda) / —_\lambda (\varphi * \pi_\lambda) = S_{ij}/\Phi_j \in [0 ; 1]$. This value is converted in percentages for an easier reading and the solution of possible border effects. The methodology is summarised by the following figure :

Figure 10: *The 3 steps of implementation of Corilis (Lacaze, 1999)*



Example of applications are many. The following map of Europe depicts the dominant land cover types defined from an aggregation of CLC in 5 classes (Urban-artificial, broad agriculture, mixed agriculture/grassland, forests, natural/semi-natural areas). In each hexagon is attributed a land cover type (a colour) when this type has a value > to its [mean value + standard deviation]. When two types are in competition (a very rare case), priority has been given to the less important type at the overall level. What remains in grey is made of areas where no dominant characteristic can be specified at this stage.

Figure 11: *CORILIS : dominant land cover types in Europe (based on [mean value + standard deviation] analysis) - source IFEN*



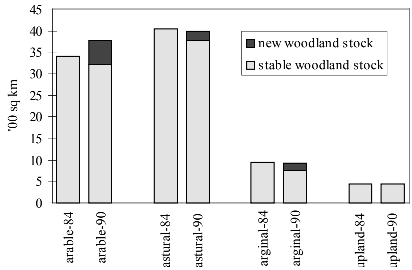
A wide range of applications, geographical as well as statistical, helps in defining the context of specific zones, as well as in improving the interpretation of indicators.

4. Indicators development

As long as indicators provide information on interactions (between economic agents, between natural phenomenons or between human activities and the environment) contextual information is essential for their interpretation.

4.1. Another example from the Countryside Survey

Figure 12: *Changes in Broadleaf Woodland Stock for the Four Major Landscape Types of GB, 1984-90 (Haines-Young, 1997).*



Note: Each pair of bars represents the stock of woodland at the start and end of the survey period for a different landscape zone; thus arable -84 and arable -90 are woodland cover in the arable zone in 1984 and 1990 respectively.

Example of the specific role of landscape typology has been given previously with the presentation of the Countryside Survey of Great Britain. The figure below illuminates the value added by integrating indicators (here, the growth of forests) with information on major landscape types.

To a large extent, what has been said previously about stratification of samples can be mirrored in terms of interpretation of the indicators, as long as the issues for which they are established are correlated with geographical factors such as climate and relief, vegetation coverage or pressure by human activities. When zoning improves the local relevance of the indicators, a typology of zones permits some statistical extrapolation of the information that they provide.

4.2. Smoothed data for indicators analysis : LaCoast & Corilis

Such an approach has recently started to be tested with the LaCoast database developed by the Joint Research Centre. This database covers a strip of 10 km of the European coasts and analyses the changes in land cover between 1975 and circa 1990 (according to the availability of the data). The inventory starts from the original CLC, realised around year 1990 and looks backward for interpreting the changes in land cover, using 1995 Landsat satellite images. The first statistical analysis presented a clear confirmation of the urbanisation process at work on the coastal zone, when map showed that this process was unevenly distributed.

In 1999/2000, the Blue Plan Centre of the Mediterranean Action Plan decided to initiate a feasibility study of indicator reporting based on LaCoast. Geographers analysed the database and concluded that, despite its limited spatial resolution, it was relevant for the purpose. Their conclusions were based on the analysis of four sectors of the French Mediterranean coast, characterised by different conditions were land cover changes and in particular the urbanisation developed according to factors such as the initial importance of the urban fabric, agriculture or the existence of natural areas (possibly protected). Their comments on the changes as detected through LaCoast proved to have a more analytical content than the first statistics computed so far at the National and even the Regional (NUTS2) levels. At the same time, similarity was found between the empirical zoning made for the sake of the study and the picture given by Corilis (see the previous section) of the coastal strip.

A preliminary study by Ifen in the context of the European Topic Centre on land cover of the EEA lead to test alternative zonings based on an aggregated version of Corilis *(Uhel, Maïna, Ifen, 2000)*.

One version just keeps 3 dominant types : artificial, agriculture, forests & semi-natural areas. In each hexagon, artificial/urban is selected if its relative value comes first or second, the two other types when they come first. The over-representation of the artificial/urban type aims at taking into account the fact that the pressure generated by this type is in proportion higher than agriculture.
Another test is based on the relative values of 5 aggregated types in terms of mean or mean + standard deviation and proposes a derived classification of the dominant character of the landscape :

- highly artificial areas (split in "without any secondary dominance" and "with semi-natural as a secondary dominance") and
- little or not at all artificial features (split in "land mainly covered by agriculture", "land mainly covered by forests and semi-natural type", "land mainly covered by agriculture and forests/semi-natural" and "land with no predominant character").

One resulting map, in this case, is presented below, left.
The overlay of LaCoast data on this zoning shows the uneven distribution of the changes (urbanisation in red) over the dominant landscape types.
This first exercise will be continued by the new European Topic Centre of the EEA, in charge of the Terrestrial Environment. Integration of other features in the zoning process of the coastal zone is foreseen.

Figure 13: *Zoning the land cover of the French Mediterranean coast (10 km strip), classification according to [mean + standard deviation] (Uhel, M., 2000)*

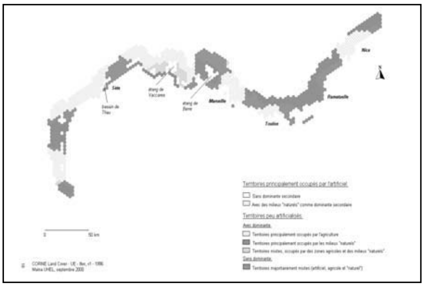
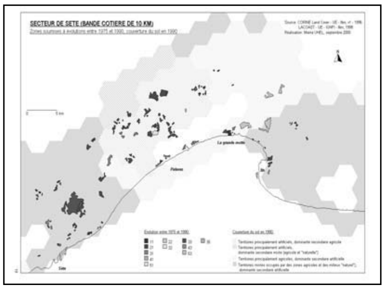


Figure 14: *Land cover changes 1975-1990 by zone and dominant landscape type (Uhel, M. 2000)*



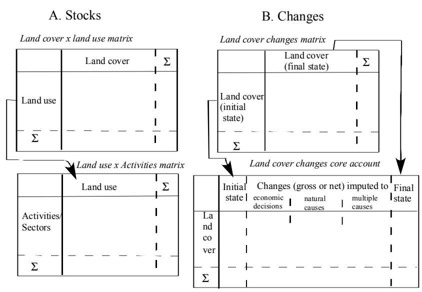
5. Accounting for land cover and land use

5.1. The new SEEA

Environmental accounts can be compiled in monetary or in physical term. They aim at providing an assessment of the natural capital in terms of stocks and flows, of its depletion due to its extraction or harvesting for production purpose and of its degradation.

In 1993, the UN published the so-called SEEA, the system of integrated economic and environmental accounts. They were constituted of a set of tables closely connected to the SNA, the system of national accounts which is the reference for most all the countries in the world. Their main purpose was to proposed adjustments of the GDP, the Gross Domestic Product which is the most prominent aggregate of the national accounts, in order to keep stock of the depletion of the natural resource and the degradation of the environment due to human activities.
In 1998, after years of experience supported by Eurostat or initiated by the countries in Europe as well as in Canada, Australia, Japan, Philippines, Indonesia, Brazil, Chile, Colombia and many other countries, a revision was decided. The process of revision is close to its end and has lead to the publication of a draft of the new SEEA in May 2001, for a final adoption by the UN Statistical Commission foreseen early next year. As compared to the previous version, the new SEEA will be a more balanced presentation of the monetary and physical accounts, in quantity and quality. The latter encompass the supply and use of the natural resource, the emission of residuals by the economy, the asset accounts of natural resource such as subsoil assets, forests, aquatic resource, water, land and the ecosystems.
The land accounts are presented according to the framework defined by a task force of the UN-ECE (1995). They are composed of core accounts and supplementary accounts. The core accounts describe the stocks and changes in land cover and land use.

Figure 15 : *Core accounts of land cover and land use (UNECE, 1995)*



Land use accounts are connected to the economic activities ton which they relate. Land cover accounts identify the changes in land cover changes in three classes : due to economic decisions, due to natural causes and due to multiple causes, including catastrophes.
The so-called supplementary accounts are targeted to specific issues. They merge land data with other statistics derived from various sources. These sources may be conventional statistics (i. e. sales of fertilisers by municipality, increase of population, road traffic, subsidies to agriculture), data from monitoring networks as well as data collected from sampling surveys. Although the sources for supplementary accounts may

differ from one country to another, some comparability is insured because of a common landscape reference. In practice, the targets generally addressed are, on the one hand, artificiality & intensity of use of land, in connection with social and economic data and accounts and, on the other hand, potentiality of land, vulnerability and biodiversity.
Therefore, core and supplementary land accounts constitute a relevant framework for describing the interactions between the economy and the environment in relation to their spatial distribution.
Pilot land account have been tested in Germany, France and in Great Britain. Example are given in a report for Eurostat (Weber, Cour, Tourneux, Haines-Young, Krack-Roberg, & Schäfer, 1997).

5.2. Land cover accounts in Great Britain

In the context of the Countryside Survey, accounts for nature and for land cover changes have been compiled as a way of structuring the overall set of indicators in order to improve their interpretation. Not only the net changes are recorded but also the factors which have lead to these changes in land cover.

Table 2: *Account for Semi-Natural Cover Types in GB (after Haines-Young et al. 1996)*

'00 sq km	**1984 Stock**	semi-natural change	intensification	extensification	afforestation	deforestation	development	reclamation	**1990 Stock**	**net change**	se of change	% change	loss	**% 1984 stock lost**
non-agricultually improved grass	**15.1**	0.9	-3.1	5.3	-0.3	0.1	0.0	0.1	**18.1**	**3.0**	2.4	19.6	4.0	**26.5**
upland grass	**64.2**	-2.6	-5.9	6.5	-1.4	0.5	-0.2	0.0	**61.2**	**-3.0**	3.2	-4.6	19.2	**29.9**
dense bracken	**41.6**	-5.2	-0.1	1.2	-0.7	0.9	0.0	0.1	**36.9**	**-4.7**	4.1	-11.4	12.2	**29.4**
purple moor grass-dominated moorland	**39.0**	0.3	-0.1	0.2	-2.4	0.3	0.0	0.0	**37.2**	**-1.8**	1.6	-4.5	6.1	**15.6**
moorland grass	**86.1**	-1.4	-0.7	0.2	-1.0	0.0	-0.3	0.0	**82.9**	**-3.2**	2.0	-3.7	11.7	**13.6**
unmanaged grassland and tall herb	**22.3**	1.0	-4.7	11.9	-1.1	0.9	-2.4	0.8	**28.7**	**6.5**	2.6	29.0	9.6	**42.9**
dense heath	**46.4**	0.8	-0.1	0.1	-0.4	0.1	0.0	0.0	**46.9**	**0.4**	1.4	0.9	4.4	**9.5**
drier northern bogs	**49.5**	-0.3	-0.2	0.0	-3.2	0.0	0.0	0.0	**45.9**	**-3.6**	3.2	-7.3	5.3	**10.6**
Total semi-natural	**693.7**	-0.7	-18.1	29.1	-13.2	6.2	-3.2	1.3	**695.1**	**1.3**	**1.3**	**0.2**	**90.2**	**13.0**

"The deeper insight into the implications of land cover change provided by the accounting approach can be further illustrated by the transformations associated with the semi-natural cover types recorded for GB by Countryside Survey 1990 (Table 2). Overall, these data show the total area of semi-natural land to be stable, but that this stability masks a considerable turnover of land within the category. The columns for 'intensification' and 'extensification' are particularly interesting. The former records the land that has been lost from a semi-natural type to arable or an intensively management grassland cover class; extensification records change in the opposite direction. If one looks at net change alone, then they might be interpreted as consistent with the goals of sustainability. By contrast, if one considers the condition of the resource in relation to the turnover of land and the exchange of land via intensification and extensification, then a very different conclusion can be drawn. Only if the land gained was of equivalent ecological value to that lost, could the stability in area be considered consistent with the goals of sustainability." (Haines-Young, 1997)

5.3. Land cover accounts by ecozones and types in Franche-Comté, France

In the test realised in Franche-Comté, the analysis of land cover change has been carried out from the comparison of two Landsat TM satellite images for 1984 and 1992. The purpose was to produce a retrospective view of the land cover and changes in the area, the 1992 situation being established by the Corine land cover national (and European) map.

The methodology for producing the retrospective map consisted in four steps. In the first one, the two Landsat images have been merged and processed in order to identify and map the changes which occurred. The main compound analysis was made around the two axes of « brightness » and « greening ». Then, in a second step, the type of change was specified according to the Corine land cover classification. This was done by visual photo-interpretation. Due to the first step, photo-interpretation was restricted to the changing areas only.

The statistical results of the analytical phase of the study can be expressed in the following transition matrix, referring to an aggregation of CORINE land cover classes :

Figure 16: *Matrix of changes of land cover in Franche-Comté, 1984-1992*

Land abandonment

CLASSES	111	112	12,3,4	21+22	231	242	243	31	32	4+5	*TOT 84*
111	**273**										*273*
112		**11367**									*11367*
12+13+14			**2435**								*2435*
21+22				**39394**	163	941	148	9			*40655*
231		60	22	1035	**65029**	1386	2926	10	68		*70536*
242		31	189	5151	1118	**40227**	690	52	31		*47489*
243		115	51	63	76	73	**13830**	346	2599		*17153*
31		22	92					**141098**			*141212*
32		12					161	203	**3249**		*3625*
4+5										**1906**	*1906*
TOT 92	*273*	*11607*	*2789*	*45643*	*66386*	*42627*	*17755*	*141718*	*5947*	*1906*	***336651***
92 - 84	0	240	354	4988	-4150	-4862	602	506	2322	0	0

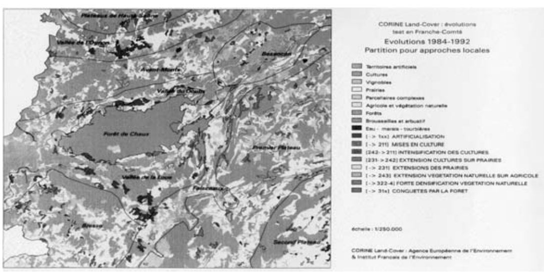
Urbanisation Intensification of agriculture

Aggregated CLC classification as referred in the Franche-Comté research:
111 : Dense urban fabric / 112 : Less dense urban fabric / 12,13,14 : Industrial, trade, transport and leisure areas/
21+22 : Arable land and permanent crops (smaller unit > 25 ha)/
231 : Pastures/ 242 : Mixed agriculture (parcels < 25 ha)/ 243 : Agriculture mixed with natural features (parcels < 25 ha)
31 : Forests / 32 : Semi-natural and natural land/ 4 : Wetlands / 5 : Water bodies

At this stage, some interpretation of the changes is possible. In most cases, changes involve the so-called mixed classes of CLC, such as in the cases of urban sprawling, intensification of agriculture, or land abandonment by agriculture. Changes take place, in many occasions, as an extension of a mixed class over an homogenous one (112 over 21 in the first stage of urbanisation or 111 over 112 in a second stage; 21 over 242 or 243 in case of intensification of agriculture; 31 over 243 in the case of land abandonment). Some interpretation of the changes is therefore possible (and not only a comment on the + & -).

The third step was the landscape analysis in homogenous environmental regions (valleys, plateaux, urban areas) which was realised quickly at this stage. A second set of statistics resulted from the overlaying of this map with the map of land cover changes.

Figure 17: *Map of land cover changes in Franche-Comté, zonal approach*



Finally, Land cover accounts (core accounts in the sense of the overall Land accounts framework) have been realised (Table below). They demonstrate the importance and location of phenomenon such as intensification of agriculture (in valleys), land abandonment (on plateaux) and urbanisation (around a major town). Even at this preliminary stage, Land accounts prove to be a help for land planning and environmental assessment and policy making.

Table 3: *LAND ACCOUNTS / Franche-Comté Test Area/ 1984-1992*
Core Accounts : Zonal Account (overall table) - Unisfere & Ifen, 1997

Ecozones/geosystems	Besançon	Valleys	Plateau of Haute-Saône	Avant-Monts	Forest of Chaux	Bresse	Faisceau	"1rst" Plateau	"2d" Plateau	TOTAL
Land Cover 1984										
111- Continuous urban fabric	206	67								273
112 - Discontinuous urban fabric	3000	4559	63	808		704	998	904	331	11367
12,13,14 - Other artificial surfaces	1154	1045		59			177			2435
21,22 - Arable land and permanent crops	341	21783	3621	7885		2622	2413	1990		40655
23 - Pastures	1649	15832	1235	6061	52	12816	6252	17511	9128	70536
242 - Complex cultivation patterns	4485	13430	707	6478	236	4311	5088	12347	407	47489
243 - Land with agriculture and natural vegetation	926	4684	367	1610		1991	2328	4696	551	17153
31- Forests	5791	19617	3245	15402	23023	15898	8495	38328	11413	141212
32 - Schub, herbaceous vegetation associations	561	456		155		45	489	853	1066	3625
4 - Wetlands							67	97	299	463
5 - Water bodies	315	903		15		190	20			1443
TOTAL	18428	82376	9238	38473	23311	38577	26327	76726	23195	336651
Changes due to economic decisions	***375***	***6582***	***581***	***620***		***1562***	***657***	***692***	***90***	***11159***
A - Agriculture										
A1 - Specialisation/intensification of agriculture	29	5210	433	199		804	598	243		7516
A2 - Other agriculture moves	9	1142	100	395		558	31	255		2490
B - Urbanisation	337	153				8	28	42	26	594
C - Other anthropogenic causes		77	48	26		192		152	64	559
Changes due natural and multiple causes	***14***	***613***	***75***	***423***		***336***	***1035***	***3745***	***539***	***6780***
D - Natural increase or depletion	14	501	75	409		260	1035	3745	539	6578
E - Catastrophic losses										
F- Multiple causes		112		14		76				202
Total changes (-)	***-389***	***-7195***	***-656***	***-1043***		***-1898***	***-1692***	***-4437***	***-629***	***-17939***
Land Cover 1992										
111- Continuous urban fabric	206	67								273
112 - Discontinuous urban fabric	3179	4579	63	808		712	1005	904	357	11607
12,13,14 - Other artificial surfaces	1312	1178		59			198	42		2789
21,22 - Arable land and permanent crops	361	26022	3940	7843		2699	2919	1859		45643
23 - Pastures	1609	14902	897	5814	52	12680	5655	15906	8871	66386
242 - Complex cultivation patterns	4327	9661	791	6459	236	4286	4542	11918	407	42627
243 - Land with agriculture and natural vegetation	819	4820	205	1777		1672	2597	5349	516	17755
31- Forests	5737	19643	3293	15428	23023	16090	8495	38377	11434	141520
32 - Schub, herbaceous vegetation associations	563	601	49	270		248	829	2274	1311	1443
4 - Wetlands							67	97	299	1443
5 - Water bodies	315	903		15		190	20			1443
TOTAL	18428	82376	9238	38473	23311	38577	26327	76726	23195	336651

Conclusion

Land in Europe is a scarce natural resource. However, land cover as such is not an economic asset or an ecological patrimony. Instead, land cover is the resulting track, the foot-print, of the many elements which are interacting. Land cover is, therefore, a key component of the analysis of the relation between the environment and the human activities. Of course, a deeper understanding of the processes at work requires supplementing land cover information with quantitative and qualitative data on the economy as well as on the natural environment. This is facilitated by the development of the data base and tools for data handling and analysis such as the Geographic information systems and the geo-statistical methodologies.

References

Barr, C.J. Bunce, R.G.H. Clarke, R.T., Fuller, R.M., Fruse, M.T. Gillespie, M.K., Groom, G.B, Hallam, C. J., Hornung, M., Howard, D.C and Ness, M.J., (1993) *Countryside Survey 1990, Main Report.* Countryside Survey 1990 Series, Volume 2, DoE, London.

Chataignon Ph. & Grasland C. (1998) in "Séminaire de géostatistique – session 1997 – 1998" INSEE n° H9802, Paris.

Countryside Survey 2000 (2000), http://www.cs2000.org.uk

DETR (2000), Accounting for nature : assessing habitats in the UK countryside, Department of the Environment, Transport and the Regions : 134

Gallego, F.J., (1999), Crop Area Estimation in the MARS Project, Conference on ten years of the MARS Project, Brussels, April 1999, JRC: 11

Gallego, F.J., Carfagna, E., Peedell, S., (2000). The use of CORINE Land Cover to improve area frame survey estimates in Spain. Research in Official Statistics: 19

Haines-Young, R.H., Watkins C., Bunce RGH and Hallam CJ. (1996), Environmental accounts for land cover, Countryside 1990 Series N°8, Department of the Environment, London

Haines-Young, R.H., (1997), UK : Land Cover Accounts., in Weber, Haines-Young et .

Haines-Young, R.H. (1999), Environmental accounts for land cover : their contribution to state of the environment reporting. Transactions of the Institute of British Geographer 24 : 441-456

Lacaze, M. (1999), Géostatistique « Grands traits de l'occupation des terres », internal document, Ifen

Lacaze, M ; et al. (2000), Ces terres qui nous entourent, Les données de l'environnement n° 51, Ifen, Orléans, France. http//:www.ifen.fr

Leonard, J. and Crouzet, P. (1999). Contribution to EUROWATERNET. River water quality in France. National prototype (phase 1), Ifen & Office International de l'eau : 40.

UN-ECE, Conference of European Statisticians (1995) : Physical Environmental Accounting : Land Use/ Land Cover, Nutrients and the Environment, Etudes et Travaux n°4, Ifen

Weber, J.-L., Cour, Ph., Tourneux, F.-P., Haines-Young, R.H., Krack-Roberg, E., & Schäfer, D. (1997) Emission Structure Information System : Physical Accounts for Land cover /Land use and related Changes in Artificiality of Land and Biodiversity, Report for Eurostat, Ifen: 145

Electronic Mapping Utility
A Solution to Geographic Process

Gordon Reichert, Antoine Jabourri,
Sylvain Cloutier, Paul Grant

Agriculture Division, Statistics Canada, 12-D2 Jean Talon Building, Ottawa, Ontario,
K1A 0T6, Canada, Tel.: +1 - 613 951 3872
e-mail: gordon.reichert@statcan.ca

Abstract: This paper describes the Electronic Mapping Utility developed, tested and implemented with the Census of Agriculture by Spatial Analysis and Geomatics Applications, Agriculture Division, Statistics Canada. The Electronic Mapping Utility was developed to eliminate the bottleneck experienced during past Censuses of Agriculture when paper maps were used to complete the manual edit to verify and assign the farm residence and the farm headquarters locations. Visual Basic and MapObjects were used to develop the application within a Geographic Information System framework. Approximately 42,000 records or 15 per cent of the records that require manual editing will be processed using the Electronic Mapping Utility. The Electronic Mapping Utility is expected to reduce the editing time by 300 person days.

Keywords: Census of Agriculture, Electronic Mapping Utility, Geographic Edit, Geographic Information System, GIS, MapObjects, Visual Basic.

1. Introduction

During previous Censuses of Agriculture, Statistics Canada used paper maps to locate and assign missing, or correct invalid, farm headquarters geography in its Geographic Edit process, a practice that has been in place for many years. Although paper maps permitted the manual editing to be successfully completed, they were identified as a major bottleneck in the editing process. In 1996, approximately 42,000, or 15 per cent of the 280,000 questionnaires collected by the Census of Agriculture were manually processed using paper maps to locate and assign missing, or correct invalid, farm headquarters geography. Paper maps, however, resulted in handling delays and it was evident that an efficient solution was needed to resolve the manual edit bottleneck. To expedite the manual editing process for the 2001 Census of Agriculture, the Agriculture Division developed the Electronic Mapping Utility. As the name implies, the application is electronic thereby permitting unlimited, simultaneous access to the databases, digital boundary files, questionnaire images and scanned maps. The Electronic Mapping Utility is expected to reduce the manual editing time by 300 person days.

To better appreciate the benefits of Electronic Mapping Utility, one must have an understanding of the Geographic Edit process of which the Electronic Mapping Utility is an integral component.

2. Geographic Edit

The mandate of the Geographic Edit process is three-fold:

- To ensure that the resident geographic identifiers are valid
- To assign or verify farm headquarters geographic identifiers for each Census of Agriculture questionnaire, and
- To verify/edit units of measurement (arpents) for farmland in the province of Quebec. The arpent is equal to about 5/6 of an acre and is used only in some localities within the province of Quebec. In order to accept or assign a correct unit of measure the correct farm headquarters geography must be available.

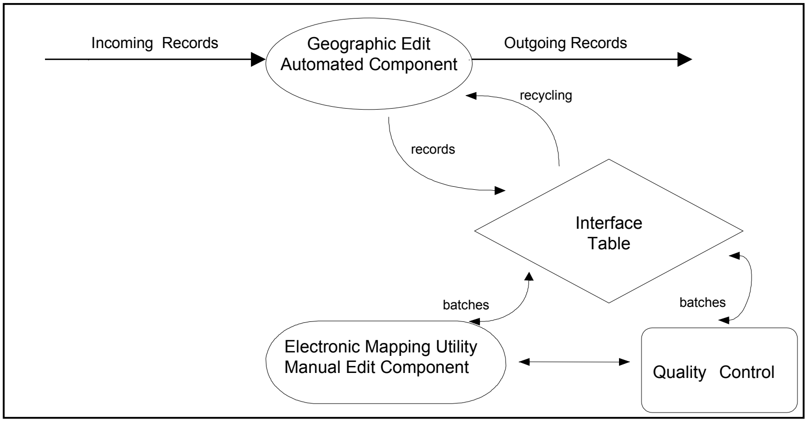
The most common causes that result in a questionnaire failing one or more edits are:

- Farm is large and information assigned must be verified
- Farm headquarters geography is located within an 'Urban' area
- No entry for residence geography or farm headquarters geography
- Incomplete residence geography
- Incorrect residence and farm headquarters geography captured
- Residence geography or farm headquarters geography is not listed on the Agriculture Geographic Master File.

It is imperative that all records contain the correct farm headquarters geographic variables since all further processing and publications are based on the farm-headquarters geography. The farm residence is also important for Agricultural Population linkage.
The residence and farm headquarters location are identical in 85 per cent of the questionnaires since the farm operator's residence is located on the land farmed. However, in 15 per cent of the records the location of the residence and the farm headquarters are different. For example, a farm operator may live in a small town but the farm buildings may be located 20 kilometres from the residence. Other farm operators may live in another region of Canada several thousand kilometres from the land they own but rent the land to another farm operator. Therefore, to flag the records where the resident geography differs from the farm headquarters the record is processed through Geographic Edit which has both an automated component and a manual component (Grant et al. 2001).

Automated Component: All questionnaires processed during the Census of Agriculture are routed through the automatic component of Geographic Edit (Figure 1). In the automated component, the residence geography is verified against the Agriculture Geographic Master File to ensure that the residence geography is valid. Once the validity of the residence geography has been confirmed, Geographic Edit attempts to automatically assign a farm headquarters geography. This is done using information from the Census of Agriculture questionnaire or information stored on the Farm Register. Records that are successfully assigned a farm headquarters geographic identifier and do not fail any other edits within the automated function proceed via outgoing records to the next step within the Central Processing System (CPS). Records failing one or more of the 15 edit criteria will be diverted from the regular CPS processing stream and grouped into batches. Individual batches are assigned to a single user, contain a maximum of 30 edit records and are grouped by province. The batches are then sent to the Interface table which serves as a temporary repository of flagged records held for manual processing.

Figure 1. *Components of Geographic Edit with Quality Control*



Manual Component: The manual edit portion of the Geographic Edit process requires staff to complete two main tasks: first, locate and assign farm headquarters when the automated component could not do so and second, for large farms, verify that the information assigned automatically is correct. Verification is important since large farms have a significant impact on data associated with a dissemination area. Often data users are interested in historical comparisons with previous Census data thus making the correct location of farm headquarters essential.

The manual process is made up of two parts: a data processing component and a geographic component. The data processing component transfers the Census of Agriculture questionnaire images (Jones and Green, 1998) and the geographic data present on the CPS to the Electronic Mapping Utility. In turn, when a user identifies a location or geography for a residence or farm headquarters, the Electronic Mapping Utility will return to the CPS the geographic data identified. The final location of the farm headquarters requiring manual intervention will be saved as a point in a Latitude/Longitude format and will retain all of the geographic boundary codes used by the Census of Agriculture.

3. Quality Control

The Geographic Edit production environment provides a convenient framework for implementing a Statistical Quality Control (QC) operation. Although the QC operation targets the work undertaken by the Electronic Mapping Utility staff, it remains independent of the Electronic Mapping Utility. The on-line correction procedures have on-line quality assurance, quality control, quality check and support procedures (Sampson 2001).

4. Electronic Mapping Utility

System Requirements: The Electronic Mapping Utility was developed to run on systems using Windows NT 4.0. Each computer has a processor speed in excess of 450Mhz. A minimum of 128Mb of Random Access Memory and 512 Kb of cache. A computer screen size of 19" is sufficient but a 21" monitor is recommended with a minimum resolution of 1280x1024 pixels and 64K colour depth.

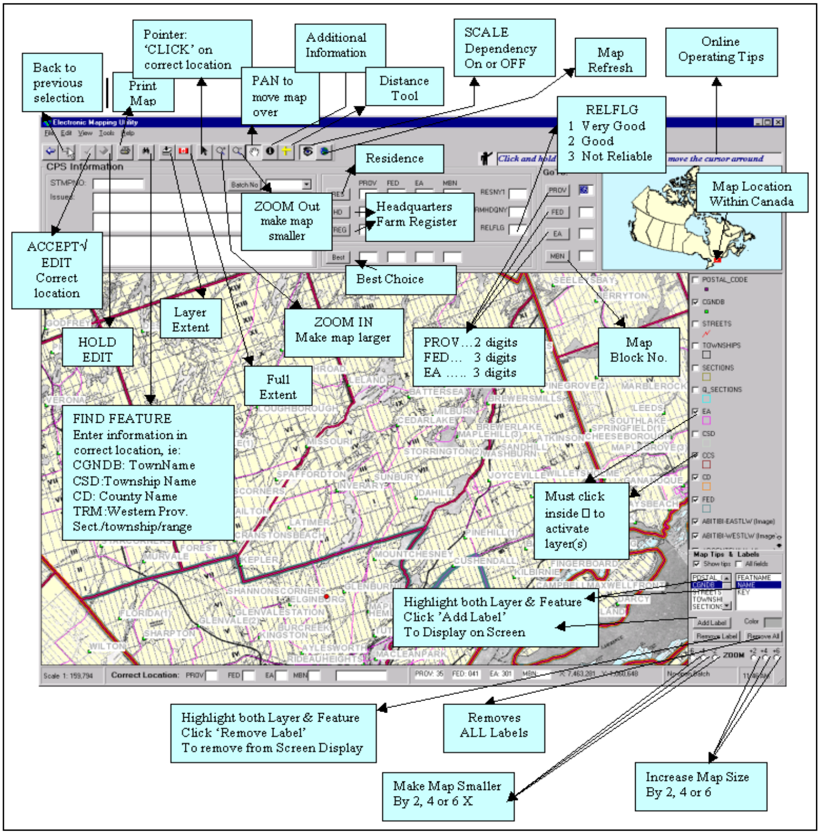
Software Requirements: The CPS is based on an Oracle 8.1 database. All of the Census of Agriculture computers utilize Oracle*Forms 6.0 and Oracle Net*8. FileNET's Panagon image viewer is also installed on all of the computers. The Electronic Mapping Utility runs simultaneously with Oracle*Forms, Oracle Net*8, Panagon viewer and can extract and update data in the Oracle 8.1 database remote on the server. To access and serve the digital boundary files and scanned maps requires each computer to have MapObjects installed.

Graphical User Interface: The Electronic Mapping Utility accesses the Census of Agriculture databases and digital boundary files (Table 1) via a custom-built Graphical User Interface (GUI) thereby eliminating the need for paper maps. A robust, but easy to use GUI (Figure 2) has been specifically developed using Visual Basic and MapObjects within a Geographic Information System (GIS) framework to simplify the editing task within Geographic Edit.

Table 1. *Geographic Files of the Canadian Provinces and Territories.*

Geographic File Description for the 2001 Census of Agriculture	
PROV	Province/Territory
FED	Federal Electoral Districts
CD	Census Divisions
CCS	Census Consolidated Subdivisions
CSD	Census Subdivisions
CMA/CA	Census Metropolitan Areas/Census Agglomerations
CCD	Census Commissioner Districts
CT	Census Tracts
Collection EA	Collection Enumeration Areas
Collection block	Collection Blocks
Dissemination Area	Dissemination Areas
Dissemination block	Dissemination Blocks
Hydrology	Lakes, Rivers and Streams
Roads	Street Network File
Towns, municipalities, villages, small towns	Place Name Master File
Postal Code File	Postal code file queriable by EA

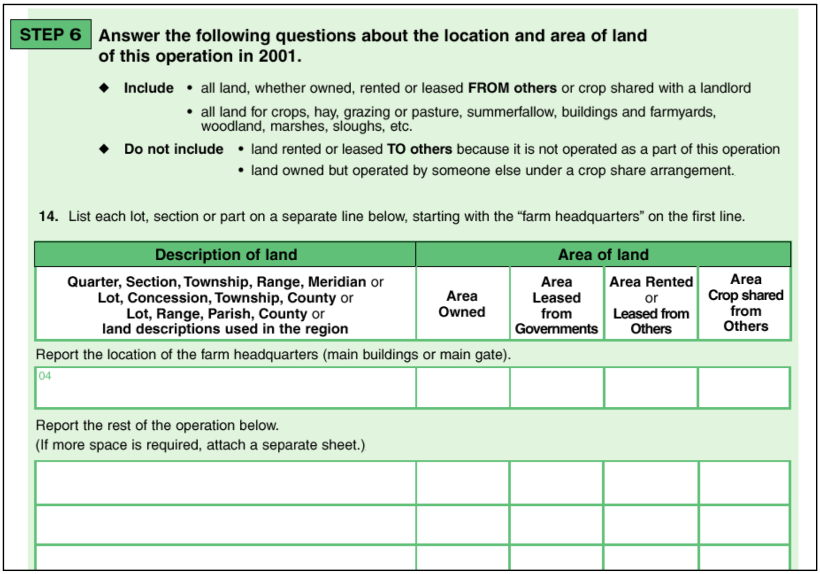
Figure 2. *Graphical User Interface for the Electronic Mapping Utility*



Functionality: Standard GIS functions such as zoom in, zoom out, pan, search, save, print, et cetera are part of the Electronic Mapping Utility. For a user, the Electronic Mapping Utility provides a single screen that includes all necessary options required to successfully complete an edit. The application is protected with a username and password requirement. It is only with proper security clearance that a user will be granted access to the batch jobs within the Oracle database on the CPS. The batch jobs contain the records that have failed the automated component of Geographic Edit. To help the user resolve the manual edit a "best" starting point is assigned within the automated component of Geographic Edit based on the failed edits. If a best starting point could be determined, in part or in whole, each best start field will contain a value; otherwise the best start field will be set to NULL. When the user activates a batch job the Electronic Mapping Utility reads the best starting point and the software automatically zooms in to the best starting point region on the map. If the best starting point is NULL the Electronic Mapping Utility will default to a province level view at which point the user must determine the correct geographic location from additional information such as the questionnaire images.

Questionnaire images: Information about the automated edit failure is passed on to the GUI along with the image of the questionnaire (Figure 3). The questionnaire images are stored on the CPS server and are passed on to the GUI via FileNet. Resolving the automated edit error requires the user to review the information provided by the respondent. Experience has proven that the land description helps resolve most of the edit errors, therefore, the default image loaded in the GUI is the land description portion of the 16 page Census of Agriculture questionnaire. If required, the user can view any portion of the questionnaire image via the GUI. General or specific geographic information, retrieved from the questionnaire image, can be quickly located by entering the information into the search tool.

Figure 3. *Land Description from the Census of Agriculture Questionnaire.*



STEP 6 **Answer the following questions about the location and area of land of this operation in 2001.**

◆ **Include**
- all land, whether owned, rented or leased **FROM others** or crop shared with a landlord
- all land for crops, hay, grazing or pasture, summerfallow, buildings and farmyards, woodland, marshes, sloughs, etc.

◆ **Do not include**
- land rented or leased **TO others** because it is not operated as a part of this operation
- land owned but operated by someone else under a crop share arrangement.

14. List each lot, section or part on a separate line below, starting with the "farm headquarters" on the first line.

Description of land	Area of land			
Quarter, Section, Township, Range, Meridian or Lot, Concession, Township, County or Lot, Range, Parish, County or land descriptions used in the region	Area Owned	Area Leased from Governments	Area Rented or Leased from Others	Area Crop shared from Others

Report the location of the farm headquarters (main buildings or main gate).

04				

Report the rest of the operation below.
(If more space is required, attach a separate sheet.)

Search Tool: A dynamic search tool (Figure 4) has been included in the GUI to assist the user in locating an area described in the census questionnaire. Some of the most common descriptions provided by the respondent are: county names, municipalities, concession and district lot numbers, parish and range names, lakes and rivers, town names, section and quarter section, township/range/meridian, highways and roads, street names, latitude and longitude coordinates and postal codes. This tool returns the search result to the user. Selecting the appropriate record from the search results allows the software to quickly locate the area of interest and load the map for the user to cross-reference the location with other supporting data such as the geographic boundary files (Figure 5).

Figure 4. *Electronic Mapping Utility Search Tool.*

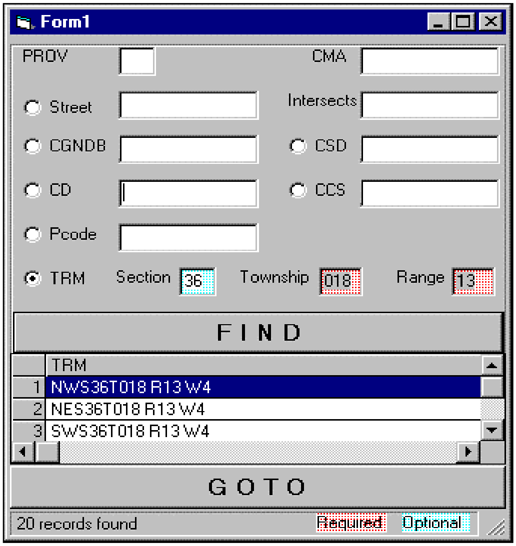
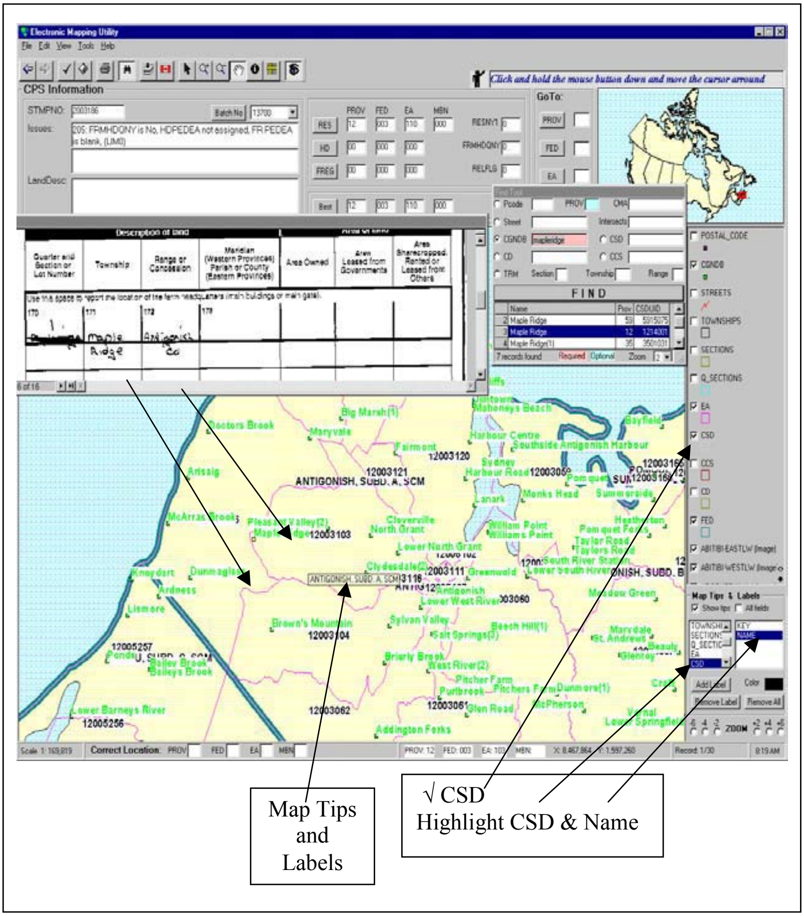


Figure 5. *Farm Headquarters Selection Process Using the Electronic Mapping Utility.*



Geographic Boundary Files: The user can activate or de-activate various geographic boundary layers depending on the level of supporting information required. In addition to the digital boundary files developed and maintained by Statistics Canada, 200 paper maps were scanned and saved as raster files and geo-referenced to the shapefiles providing an additional level of information to assist the users in determining the correct location of the farm headquarters.

Location of Farm Headquarters: Once the user identifies the correct location of the farm headquarters they can point and click on the GUI map with the mouse to automatically assign the selected geographic location to the "Correct Location" box on the GUI. The user can save the edit to the Interface table by activating the save button on the GUI. Immediately after accepting the edit the next record within the edit batch is automatically loaded and the edit process is repeated.

Putting an Edit on Hold: If necessary an edit can be put on hold and re-activated at a later time. Suspending an edit is generally done when a user requires the assistance of a supervisor to resolve an edit selection. Putting an edit on hold does not stop the edit process, it only suspends the edit of the record indicated. A user can continue to process other records within the batch but the batch edit results will not be passed on to the Interface table until all records within the batch have been resolved. More than one record within the batch can be put on hold. If the supervisor is unavailable and all other records within the batch have been resolved the user can activate another edit batch. When the supervisor is able to provide assistance the edit on hold can be re-activated. Re-activating an edit on hold reinstates the edit exactly where it was suspended. Once the edit has been resolved the batch can be saved and the batch edit results are then saved to the Interface table.

Repeat as necessary: Records which have failed at least one edit in the automated component of Geographic Edit and have been manually resolved using the Electronic Mapping Utility will be re-processed through the automatic component of Geographic Edit. Recycled records will continue to be re-processed until they no longer fail any of the automated edit criteria within Geographic Edit.

5. Future Direction

Based on test results and user response, the Electronic Mapping Utility will be modified for other geo-spatial applications useful to the Census of Agriculture such as "Data Validation" and "Geographic Splits". Other Divisions within Statistics Canada have also evaluated the Electronic Mapping Utility and have requested a modified version specific to their criteria.
Efforts were made to build the utility in order that it could form the basis for similar applications on other surveys in Canada or in other countries. A modular approach was used in developing the Electronic Mapping Utility to ensure that modifications can be completed quickly and easily. As well, the application relies on inexpensive computer hardware and software already available in the government offices of most countries.

6. Conclusions

The Electronic Mapping Utility has been developed to eliminate the bottleneck experienced when paper maps are used to complete the manual edit to verify and assign the farm residence and the farm headquarters location. The Electronic Mapping Utility was designed to provide an efficient and cost-effective solution for conducting manual edits. These objectives have been successfully met. Electronic delivery of the application has greatly increased the ease and flexibility of the application, provided greater reliability and efficiency in operation and easily accommodated new and growing processing requirements. The modular design philosophy and the close interaction between supervisors, software programmers and system operators have been shown to be the essential ingredients to a successful application.

References

Grant P., Cloutier S., Graham N., Olsheskie, S. (2001) Geographic Edit Reference Manual, *Internal Report*, Agriculture Division, Statistics Canada. 1-129.

Jones M., Green, I. (1998) Utilization of Document Imaging Technology by the 1996 Canadian Census of Agriculture, *Proceedings of Agricultural Statistics 2000*, Washington, D.C., 202-219

Sampson G. (2001) Census of Agriculture 2001 Quality Control Methodology Specifications for Manual Geo Edit, *Internal Report,* Business Survey Methods Division, Statistics Canada. 1-20.

Lower Census Costs Through the Use of Administrative Sources

Gerry Brady
Central Statistics Office, Skehard Road, Cork, Ireland
Tel.: +3521 - 359 154
e-mail: Gerry.Brady@cso.ie

Abstract: The CSO recommended to the Irish Government and the National Statistics Board that the possibility of conducting a Census of Agriculture (COA) in 2000 by post using administrative farm registers was considered feasible. The previous Census in 1991 was conducted by interview immediately after the 1991 Census of Population. An administrative farm registers based postal COA was expected to cost around £1.3m. This was estimated to be £2m to £2.5m below what an interviewer based Census would have cost. The Government approved the holding of a Census of Agriculture based on this strategy. This Paper reviews how and why the CSO is making increased use of farm administrative information in conducting statistical farm surveys. Difficulties encountered are also discussed.

1. Old farm surveys methodology

A Census of Agriculture (COA) is conducted in Ireland approximately every ten years. Results from the 2000 Census are expected to show that there are currently around 140,000 farms compared to 171,000 in the 1991 Census of Agriculture. Between Censuses, two annual sample surveys of around 25,000-30,000 are conducted each June and December. There has been substantial change in the data collection methodology for both the Census and the sample surveys over the past fifteen years. The pattern of change has been to make increased use of existing administrative information, to reduce the response burden on farmers, to have access to an up-to-date register of farmers and to reduce data collection costs.

Up to 1987-1988 all CSO farm surveys were conducted by personal interview. Temporary enumerators were recruited at local level before each survey. Each enumerator was required to interview every farmer within a particular geographical district. The local area districts accounted for one-quarter of the country and they were annually re-surveyed each June and December. The recruitment and supervision of these interviewers was undertaken by the local Police force (Gardai). The system became untenable for a variety of reasons: cost; unwillingness of the Gardai to continue involvement given other priorities; informality of the field supervisory and enumerator procedures and rigidity of the sampling areas. The requirement to only interview farmers in a fixed set of districts meant that farmers in other districts were never sampled between Censuses. The requirement to obtain complete coverage within a district resulted in an undue response burden on some farmers.

2. EU Restructuring Plan

As part of an EU restructuring Plan, postal surveys were introduced in the fixed enumerator sampling areas during 1987-1988. The enumeration process involved collecting the name and address of each farmer. As part of the transition process, these were computerised to represent a register of farmers in the traditional enumeration districts. The new methodology resulted in a more flexible sample selection (stratification by farm size) and eliminated field costs.
As part of the restructuring Plan, an interviewer based Census of Agriculture (COA) was conducted in 1991. The Census was conducted immediately after the Census of Population (COP). The primary farm register used for the COA was the list of households and premises identified during the COP field-work as being farm households. This strategy reduced Census costs through the use of the COP field force, maps etc. The Agriculture Division was thus saved the costly and time-consuming tasks of recruitment, training and enumeration area field preparation work. It also allowed us to benefit from COP expertise in conducting such operations. The 1991 Census was used to create a complete register of all farmers in the State.
In the 1992-1999 inter-Censal period, postal sample surveys covering the whole country were conducted. This worked well in relation to costs and stratification but there was a difficulty in maintaining a fully up-to-date statistical farm register. There were two main strategies for identifying changes in the farm register:
A twice-yearly analysis of changes to the main DAFRD herd register was used to identify possible births and deaths. These changes were then investigated by means of register surveys to those households in March and September. However DAFRD held details on farmers in many separate databases. Only changes in the main register were examined, as many of the other registers were not fully computerised. Changes to the main herd register could occur for a variety of reasons other than indicating a new or retired farmer.
Replies from farms included in the CSO June and December surveys were the second main strategy. The farmer was asked to identify the names of persons to whom he had recently sold or let land. These persons were often expanding an existing holding rather than becoming a new entrant into farming. In practice, a significant proportion of new births were subsequently identified as duplicates of existing farmers who were already on the register under a variant of name and address. In many rural areas, clusters of persons with similar family and first names are often found. Adding duplicates in error created a significant amount of work that subsequently had to be undone when the error was eventually identified.
The main deficiency consisted of updating changes to farms that were not on the main DAFRD file without having to include them in the sample survey. The system was able to produce statistical estimates of the number and size of farms but not up-to-date names of every current farmer. A key advantage of using administrative farm registers is that they are being updated on a weekly basis at a local area level by DAFRD staff.

3. Administrative information

The 1990s were a period when an increasing amount of farm level information was being captured in a format suitable for use by statistical offices. Computer

developments also facilitated access to the new administrative data files making what had been complex programming into more routine work and also facilitating the computerisation of previously manually maintained registers. As the decade progressed it was possible to obtain access to both administrative farm registers and to the associated data files on ewe numbers, various categories of cattle, cereal areas etc.. This meant that potentially the administrative premium data could be used to examine if an entry on the administrative farm registers was associated with premium activity and the level of such activity on the farm. This possibility will be useful to prevent the addition of non-farmers who are eligible for certain premiums. The 1990s were also a period of continuing decline in the relative importance of agriculture in the Irish economy.
The definition of a farm as used by the CSO corresponds to that used in the Eurostat farm structures survey. Administrative registers can have various purposes and hence can include names not considered farms for statistical purposes. In compiling the register for use in statistical surveys the intention is to eliminate, where possible, persons not involved in farming. Persons involved in farming but with activity levels below the EU FSS thresholds can be eliminated during the processing of the survey returns.

4. Problem of multiple administrative farm registers

The Census was designed to result in some important improvements in the use of administrative sources that would facilitate continuing the methodology for subsequent sample farm surveys after June 2000. There is not yet one complete administrative register of all farmers in Ireland. DAFRD have been working towards creating such a file but going into the Census, the CSO was faced with combining multiple registers and then using extensive name and address matching techniques to eliminate duplication. The Census presented an opportunity to evaluate the differences between the various administrative registers and to significantly move towards a single client file of all farmers. Accordingly the new methodology included the following aims:

- To encourage DAFRD to fully integrate minor farm registers into the main client file. One of the medium-term CSO aims is to have access to one administrative client file of all farmers without any need to merge registers or add on specialist farmers such as those engaged in poultry, pig or horticultural production. The main DAFRD file contains most active non-specialised farmers. However it also contains persons who are not currently active in farming. For example, the main client file contained 185,000 potentially active farms at the time of the Census whereas the CSO estimated that there 144,000 active farms in June 1999.
  The six supplementary farm registers held by DAFRD numbered 196,000 names. These were matched as much as possible against the main file and duplicates removed.

- To analyse the completeness of the DAFRD farm registers by comparing them with farm registers held by other agricultural Bodies. The purpose was to identify any gaps in the DAFRD files rather than to develop a methodology based on continuous integration of registers from different administrations. A total of around 158,000 names were compiled from various registers held outside DAFRD.

- To examine the nature of entries in the administrative registers which were not farms in the statistical definition. Examples could be persons involved in environmental and retirement programs that were not active farmers. The expected result from the 2000 COA is around 138,000 farms. The combined administrative files for use in the Census was around 192,000 names. This was significantly above the estimated 134,000 farmers who had received premiums during 1999.

- To maximise and to make routine our access to the associated premiums data files. This access would allow us to focus on any significant non-respondents to the Census and to examine the Census returns for partial non-response. In the post-Census situation, it will be a useful input into sample selection and monitoring of trends between the statistical results and the administrative figures.

- To develop practices that facilitated the CSO usage of the administrative registers and data files. Examples are for DAFRD to add a phone number variable to the DAFRD client file to assist identifying potential duplicates. The CSO independently collects phone numbers from farmers as part of the farm surveys. It is a valuable item of information in register matching for births, deaths and duplicate checking.

5. Cost differences

As mentioned earlier, the CSO estimated that savings of over £2m could be made by not holding an interviewer based Census. A Census of Agriculture conducted as a postal survey using administrative registers was expected to cost around £1.3m. The principal financial savings would result from:

- No costs associated with recruitment, supervision or payment of a field force;

- No field-force travel, mapping or co-ordination costs;

- The move to using an administrative farm register meant that the CSO did not have to continue a post-Census methodology that required maintaining a statistical farm register. This is expected to result in fewer staff being required to run the surveys on an ongoing basis.

- A reliable administrative farm register should allow for smaller sample survey sizes. One purpose of the previous sample surveys was to maintain a farm register. This resulted in a sample design that resulted in all farmers being sampled over a 5-6 years period. A consequence of this was larger samples than would otherwise have been necessary.

- More timely detailed Census results. The 1991 detailed results were published in May 1994 whereas the 2000 results are due in December 2001. The time-saving was due to two reasons: minimal register work whereas in 1991 a new register of around 170,000 names and addresses had to be keyed from the Census returns in time for the December 1991 survey; the forms returned in the first week of the Census were immediately edited. In 1991 forms were only returned in late July after an interviewer had completed all of his/her work. This time saving meant that the extra staff recruited for the COA could be released at an earlier stage.

The extra costs associated with conducting the survey by post were:

- Higher processing costs as the filtering normally done by interviewers to eliminate non-farm addresses was done in-house by examining inactive Census returns;

- Higher internal staff numbers to handle the issuing and receipt of questionnaires and reminders. The Census Plan envisaged issuing three reminders to all non-respondents plus a fourth reminder to non-respondents who appeared to be active farms based on an analysis of premium data. In practice six postal reminders were issued plus a limited phone follow-up. This meant that data collection was only completed in late November 2000.

- Extra work arising from some farmers making duplicate returns. These principally arose due to the farmer being inadvertently included twice in the Census register. Another cause was that sometimes a farmer made two Census returns when a reminder crossed with his incoming first return. These problems represented a significant amount of extra work.

6. Post-Census sample surveys

Sample farm surveys were conducted in December 2000 and June 2001. The new register procedures have been working efficiently and access to administrative data is showing signs of developing further with new data files on cattle movement monitoring systems and planned tagging of all sheep arising from disease control requirements. The June 2001 survey sample size was around 5,000 farms (15%) less than previously.
When the processing of the Census is completed, we will focus on further progress towards the single complete administrative farm register.

Cost-Effectiveness of Remote Sensing in Agricultural and Environmental Statistics

Elisabetta Carfagna
University of Bologna, Dep. of Statistics, via Belle Arti 41, 40126 Bologna, Italy
e-mail: carfagna@stat.unibo.it

Abstract: Many are the ways in which remote sensing data have been used in statistically sound projects, at the design level as well as at the estimator level, for producing agricultural and environmental statistics. We present an overview of these methods, putting in evidence advantages, disadvantages and requirements and assessing their cost-effectiveness.

Keywords: Remote sensing, Cost-effectiveness, Regression estimator, Stratification, Land cover estimates.

1. Introduction

Remote sensing is a very useful tool for producing agricultural and environmental statistics, particularly for large territories. In fact, the correlation between remote sensing data and spatial variables, like land cover and land use, makes of remote sensing data an excellent base for estimating statistical parameters of these variables. Remote sensing data are used in area frame construction and stratification, in optimisation of sample designs and in estimation process. These methods are operational, but the crucial point is their cost-effectiveness. In the nineties, many authors considered the use of remote sensing at the estimator level too expensive to be cost-effective; but in the last few years, a dramatic reduction of the price of some satellite data has changed many evaluations.

2. The use of remote sensing data as auxiliary variables in the regression estimator

The most well known use of remote sensing data in the production of agricultural and environmental statistics is as an auxiliary variable in the regression estimator. The regression correction of survey estimates is a classic technique (Cochran, 1977) to estimate the mean of a variable Y known for the n units of a sample, by using an auxiliary variable X known for the N elements of the whole population and correlated with Y. This technique is generally applied by using remote sensing data for estimating the area covered by the different land cover types in a specific geographic area. Area estimates are obtained separately for each land cover c; hence we shall drop the index c. The usual regression estimator for the proportion of the area of a certain land cover c in a region is:

$$\hat{\bar{y}}_{reg} = \hat{\bar{y}} + b(\mu_x - \bar{x}) \quad (1)$$

where $\hat{\bar{y}}$ is the ground survey estimate, $\bar{x}$ is the average proportion of pixels classified as c in the segments of the ground survey sample, μ_x is the proportion of pixels classified as c in the whole region under study, and *b* is the slope of the regression between *X* and *Y*.

When the slope of the regression line has a previously fixed value (e.g. *b* is estimated from another population or from a different sample) the regression estimator is unbiased and no assumption is required about the relation between *Y* and *X* in the finite population. If *b* is a least squares estimate, the regression estimator has a bias of order 1/n. However, for large samples, it has usually a lower variance than the regression estimator with pre-assigned *b*, unless there is a good a priori knowledge of the reasonable value of *b*. The regression estimator with preassigned *b* is safer if unbiasedness is priority, distributions are highly skewed or *X* and *Y* have a non-linear link.

The regression estimator improves the accuracy of the area estimates by adjusting the estimate of the mean $\hat{\bar{y}}$ to $\hat{\bar{y}}_{reg}$ and reducing the variance (Chhikara et al. 1986, Consorzio ITA 1987, González and Cuevas 1993, Gallego et al. 1993). The variance reduction is generally by far the main benefit, and so is used as the measure of the improvement in area estimates using the regression method. If the precision to be reached for the estimates is fixed, then the use of auxiliary variables can reduce the amount of ground data to be collected. On the contrary, if the sample size is fixed, it allows improving the precision of estimates.

3. Cost-effectiveness of the regression estimator

The crucial point is the cost effectiveness of the method: in fact, the same precision of the regression estimate can be reached by increasing the number of sample segments. Remote sensing will be economical if its cost is smaller than the cost of additional sample segments:

$$(n_1 - n)p > R, \quad (2)$$

where *n* is the original sample size, *p* is the unitary variable cost (cost of ground survey, digitisation and quality control), n_1 is the sample size that allows the ground survey estimate to reach the same precision of the regression estimate and *R* is the cost of the remote sensing part of the project (image acquisition and processing).

In the case of simple regression, if the sample size is large enough, ignoring the finite population correction, the variance of the regression estimator can be approximated by:

$$Var(\bar{y}_{reg}) = \frac{S^2}{n}(1-\rho^2), \quad (3)$$

where S^2 is the population variance and ρ is the linear correlation between Y and X. Thus, by definition of n_1:

$$\frac{S^2}{n_1} = \frac{S^2}{n}(1-\rho^2)\,; \tag{4}$$

where $\frac{S^2}{n_1}$ is the variance of the direct expansion (ground survey) estimator. Thus,

$$n_1 = n\frac{1}{(1-\rho^2)}\,; \tag{5}$$

where $\frac{1}{1-\rho^2}$ equals the ratio between the variance of the ground survey area estimate and the variance after this estimate has been corrected with the aid of satellite images and is named the relative efficiency of the regression estimator or of remote sensing (η_{reg}). Substituting equation (5) in formula (2), we get the following result:

$$\eta_{reg} > 1 + \frac{R}{np}\,; \tag{6}$$

thus, $1 + \frac{R}{np}$ is the threshold value for cost-effectiveness of the use of remote sensing data in the regression estimator.

An analysis of the relative efficiency of remote sensing using SPOT and TM satellite sensors was performed in the frame of Regional Inventory action of the MARS project (Taylor et al. 1997). Table 1 shows the area-weighted average of the relative efficiencies for each of the main crops of interest.

Table 1: *Area-weighted relative efficiencies of regression estimates by crop*

Crops	1988	1989	1990	1991	1992
Wheat	1.78	1.75	2.78	2.98	3.24
Barley	1.95	1.93	1.75	2.12	2.78
Maize	1.25	1.48	3.05	2.26	n/a
Oil seeds	1.95	3.03	2.50	2.62	n/a
Dried pulses	1.77	1.53	4.14	3.61	n/a

From Taylor et al. 1997

In 1988, the thresholds for cost-effectiveness of remote sensing, were computed for each test site and satellite sensor and are presented in table 2. A comparison of tables 1 and 2 suggests that obtained relative efficiencies are generally higher than the thresholds. The threshold values also varied with satellite sensor, due to the different prices of SPOT and TM images.

Table 2: *Threshold values of the relative efficiency of using remote sensing in 1988*

	Centre	Bayern	Emilia Romagna	Macedonia	Castilla León
SPOT	1.71	1.54		2.00	1.71
TM	1.42	1.35	1.37	1.86	1.42

From Taylor et al. 1997

These thresholds were estimated on the basis of rough information. In the early nineties, the cost structure changed quite surprisingly, in the sense that ground survey costs decreased more than remote sensing costs and this resulted into higher thresholds for cost effectiveness. Particularly, for SPOT-XS sensor, real threshold values revealed to be extremely high, due to high image cost. Therefore, the MARS project of the European Union came to the conclusion that the regression correction is feasible, unless the landscape is extremely fractionated, but the precision improvement of area estimates is not enough to make up for the prices (Meyer-Roux and Vossen, 1994, p. 41).

In the same period, the United Stated Department of Agriculture (Allen, 1990) considered that remote sensing correction of estimates was not yet cost-effective, but pilot applications were maintained in few states every year to keep the know-how and software improvement.

In contrast to this, in 1992, Giovacchini and Brunetti evaluated the Italian AGRIT project carried out every year for producing area and production estimates for the main crops (FAO, 1998). On the basis of a cost of remote sensing of 23% of the total cost versus 48% of the ground visits, they reached the conclusion that remote sensing correction was cost-effective.

In 1998, Gallego followed a different approach: he considered a mid-size region (say 50,000 km^2), in which the usual sample size is about 1,000 segments. Although the survey cost per segment varies considerably in different situations, he considered 60 Euros as an average value; hence, the total yearly cost of the ground survey was about 60,000 Eros. Then, according to Taylor *et al*, 1997, he assumed that a realistic median value for the relative efficiency of a regression estimator on classified images was around 2 for the agricultural landscape in western Europe. This can be interpreted in two ways: a) remote sensing allows to reduce half of the segments with the same precision (30,000 Euros would be saved in ground survey costs, and hence remote sensing is justified if its total cost is less than 30,000 Euros) and b) the precision of estimates can be increased, either by doubling the number of segments or by introducing a regression estimator. Remote sensing is justified if its total cost does not reach 60,000 Euros.

The number of TM images required to cover a region of 50,000 km^2 depends very much on the shape of the region; Gallego considered 5 scenes to be a reasonable number i.e. approximately 15,000 to 20,000 Euros, according to market prices in 1998. With the interpretation a), cost-effectiveness was impossible with SPOT-XS and extremely difficult with Landsat TM. Thus, the author evaluated the feasibility of the regression estimator with remote sensing images acquired in the previous years in Spain. The conclusion was that, in some areas, the inter-annual correlation is high, particularly when the time lag is an even number of years.

In case b), cost efficiency was very difficult with SPOT-XS, but reachable with Landsat TM if two main conditions were fulfilled: drastic price reduction of satellite images and improving the automation of image analysis, GIS operations and statistical procedures.

The first condition is fulfilled with the new pricing policy of Landsat TM images after launching Landsat 7 (about 400 dollars per scene delivered by American distributors). The second condition is also very important, because the use of remote sensing data still requires specific skills, although a remarkable progress has been made in recent years by the USDA and the Italian ITA consortium.

3.1. Cost effectiveness for the Italian project

The Italian project provides area estimates for main Italian crops at a regional level using an area frame with 5500 segments of 50 hectares and Landsat TM images. Supervised classification is performed with images acquired in the period November - February and April - May, for winter crops, and with images of April - May and July - first days of September, for summer crops. Thus a complete coverage of the Italian territory at three dates is acquired every year.

We have performed an analysis of relative efficiencies of regression in the Italian project, summarising relative efficiencies for the different administrative regions through a weighted average. Weights are given by the percentages of crops in the different regions. When remote sensing data haven't been used in a region because they are too cloudy, a relative efficiency equal to one has been considered. We have computed the thresholds for cost-effectiveness of remote sensing: 1.40 in 1999 and 1.24 in 2000, due to the decrease of the cost of Landsat TM 7 images.

From this analysis, it becomes evident that the regression correction of ground survey estimates is cost-effective in the Italian project. We could conclude that, if the procedures of satellite data acquisition, correction, enhancement and so on are optimised and automated in an operational project, then the use of the regression estimator with remote sensing data is cost effective. Moreover the optimisation of the procedure has an influence also on the linear correlation between remote sensing and ground data, which affects the efficiency of the regression estimator, as shown in table 3.

Table 3. *Weighted efficiencies of regression for main crops, across Italian administrative regions in 1999 and in 2000*

Crops	Area (ha) 1999	Rel. Eff. Regr. 1999 (threshold 1.40)	Area (ha) 2000	Rel. Eff. Regr. 2000 (threshold 1.24)
Durum wheat	1,796,576	1.81	1,765,337	2.05
Soft wheat	611,583	1.89	492,642	2.60
Barley	340,103	1.45	299,653	1.71
Colza	30,589	2.22	22,197	1.35
Maize	1,334,071	3.62	1,382,933	5.97
Sunflower	229,018	3.25	247,850	2.85
Soy been	253,496	3.95	254,740	8.63
Sugar beet	279,981	4.07	245,644	7.60
Tomato	101,229	1.94	87,231	5.26
Tabacco	38,963	4.66	39,263	1.81
Italy	5,015,608	2.61	4,837,490	2.97

We should take into account that, if the price of 400 dollars per image is used, the threshold becomes 1.13 for the Italian project. Actually, this price cannot be considered because American distributors don't guarantee immediate delivery of remote sensing scenes and thus don't allow performing the regression correction of ground survey estimates in a very short time.

4. Estimators based on confusion matrixes

Many authors, such as Hay, (1988, 1989) and Jupp, (1989) suggested the use of a family of calibration estimators in which the biased pixel counting estimator is corrected with the help of a confusion matrix computed on a sample:
A is the confusion matrix giving the number a_{gc} of pixels of land cover g (ground truth) in a test set classified into the land use c on a set of test pixels.
Λ is the confusion matrix with elements λ_{gc} for the whole population.
Λ_g is the column vector of ground truth totals with unknown elements $\lambda_{g+} = \sum_c \lambda_{gc}$
Λ_c is the column vector with $\lambda_{+c} = \sum_g \lambda_{gc}$ number of pixels classified into class c
A_g is the same column for the test set with elements $a_{g+} = \sum_c a_{gc}$
A_c is the column vector with $a_{+c} = \sum_g a_{gc}$ number of test pixels classified into class c
We assume that A is an unbiased estimator of Λ . This is true if the test set is a random sample of the population without geographic relationship with the set of training pixels. Then, if Π_g , Π_c , P_g and P_c are the error matrices with the proportions

$$\Pi_c(g,c) = \frac{\lambda(g,c)}{\lambda(g,+)} \text{ and } \Pi_g(g,c) = \frac{\lambda(g,c)}{\lambda(+,c)} \text{ for the population} \quad (7)$$

$$P_c(g,c) = \frac{a(g,c)}{a(g,+)} \quad \text{and} \quad P_g(g,c) = \frac{a(g,c)}{a(+,c)} \text{ for the test set} \quad (8)$$

the following identities are straightforward:

$$\Lambda_g = \Pi_g \ \Lambda_c \qquad \Lambda_c = \prod_c^{'} \Lambda_g \quad (9)$$

$$A_g = P_g \ A_c \qquad A_c = P_c^{'} \ A_g . \quad (10)$$

The aim is estimating Λ_g, the number of pixels that correspond in the whole population to the land cover g (ground truth). An image classification gives Λ_c; matrices P_g and P_c computed on a random sample P_c are unbiased estimates of Π_g and Π_c, hence the direct and inverse estimators can be derived as follows:

$$\hat{\lambda}_{dir}(g) = P_g \ \Lambda_c \qquad \hat{\lambda}_{inv}(g) = P_c^{'-1} \ \Lambda_c ; \quad (11)$$

consequently, the area estimators can be given by:

$$\hat{Y}_{dir}(g) = \frac{\hat{\lambda}_{dir}(g)}{\lambda(+,+)} D \qquad \hat{Y}_{inv}(g) = \frac{\hat{\lambda}_{inv}(g)}{\lambda(+,+)} D, \tag{12}$$

where D is the area of the study region.
The direct calibration estimator based on a confusion matrix, was used in Belgium in 1992 (Gallego 1994) and in an area in the "département" Indre et Loire, in France (Brun et al. 1992). In both cases, a SPOT-XS image was used. The comparison didn't give clear conclusions; although the regression estimator looked slightly more efficient.

5. Remote sensing images for area frame construction and stratification

Remote sensing can be a precious base for estimating parameters of spatial variables, through area frame sample designs. Nowadays, when an area frame with permanent physical boundaries is constructed, strata and counting units are drawn on remote sensing images, through a manual or computer assisted procedure (FAO 1998). Remote sensing is very useful for this application and the cost is not high. For agricultural estimates, an efficient and low cost stratification is based on percentages of agriculture, often derived from photo-interpretation of remote sensing images or of aerial photos. In remote sensing projects, the effectiveness of stratification is usually assessed by the relative efficiency, which is the ratio between the variance of the estimators without and with stratification, given by:

$$\eta_{strat} = \frac{Var(\hat{Y}_{ran})}{Var(\hat{Y}_{strat})}. \tag{13}$$

However, the same reduction in variance could also be achieved by increasing the size of the ground survey sample such that:

$$\eta_{strat} = \frac{\frac{1}{n} Var(y)}{\frac{1}{n_1} Var(y)} = \frac{n_1}{n}. \tag{14}$$

The stratification will be economical if its relative efficiency exceeds a threshold given by:

$$n(\eta_{strat} - 1)p > \tfrac{s}{t} \tag{15}$$

where s is the cost of the stratification and t the number of years that the stratification is likely to be used. A land use stratification can be used for at least 10 years; thus the cost of the stratification process can be divided by 10.
When an area frame is based on square segments without permanent physical boundaries, the area frame construction doesn't require the use of remote sensing data. Anyway, images photo-interpretation can be used for producing land cover maps that allow subdividing the population of segments into non-overlapping sub-regions in such

a way that the variation of crop area per segment within each stratum is low. Practically, a square grid is overlaid on the land cover map. The attribution of each segment to a stratum is made according to a specific criterion. The criterion of majority is generally used. In essence, the stratification is adapted to the sampling grid. Each square of the sampling grid is simply attributed to the stratum with the highest share, but part of the information in the stratification is lost. This and other criteria were tested in the European Union using CORINE land cover map (Gallego et al., 1999). CORINE Land Cover is an environment oriented database, and its nomenclature has been established according to this main orientation, thus it wasn't produced to stratify the territory for crop estimation. However, it turned out that CORINE land cover is useful to improve estimates precision. This means that, sometimes, an efficient stratification can be created with very low cost, using a land cover map produced for other purposes.
Theoretically, the land use stratification carried out for an area frame with physical boundaries is more efficient, than using square segments. The reason is that the latter are generally attributed to the different strata on the basis of majority, which results in higher variability within strata. However, results of an experiment carried out in Emilia Romagna, Italy (Taylor et al. 1997) don't support this thesis.
The relative efficiency of land use stratification is strongly influenced by the characteristics of the area on which the stratification is performed. When the area frame is devoted to the estimation of areas or production of crops, since the land use stratification cannot distinguish between the different crops, the combined use of another kind of auxiliary variable allows improving the stratification efficiency.

5.1. Relative efficiency of stratification in the Italian project

The area frame currently used in the Italian project was created in the years 1992-1994 through photo-interpretation of SPOT satellite images at a scale 1:25,000 and the usual procedure of identification of primary sampling units with permanent physical boundaries. The stratification was based on land use.
We have calculated weighted relative efficiencies of stratification for the main crops in Italy (tab. 4). For each crop, relative efficiencies in the different administrative regions (η_{region}) have been summarised using the proportions (p) of crops in the regions as weights.

$$\eta_{weighted} = \frac{\sum p\eta_{region}}{\sum p} \qquad (16)$$

Cases with crops covering less than 1 % of the study area haven't been included.
Relative efficiencies are low, except for durum wheat and maize; however, this land use stratification was created using remote sensing data acquired ten years ago. The main reason for these low relative efficiencies is the fact that one of the aims of the stratification based on the land use was the identification of non-agricultural areas, where no segments were allocated. This procedure reduces the ground survey cost, but determines an underestimation of the variance without stratification and, consequently, of the relative efficiency of stratification and can produce biased estimates (see Gallego et al. 1999).
The land use stratification based on photo-interpretation of remote sensing images created strata with different rates of permanent and annual crops; but didn't discriminate

between annual crops. Thus, in some regions, land use strata were subdivided into areas with similar association of the different crops through a cluster of municipalities, based on the percentages of main crops (derived from the census of agriculture performed in 1990). The final stratification was the result of the intersection of the two kinds of stratification. The effect of the cluster of municipalities on the efficiency of stratification was relevant also after many years, although the spatial distribution of annual crops tends to change faster than the rate of annual and permanent crops.

Table 4. *Weighted efficiencies of stratification for main crops, across Italian administrative regions in 1999 and in 2000*

Crops	Area (ha) 1999	Rel. Eff. Strat. 1999	Area (ha) 2000	Rel. Eff. Strat. 2000
Durum wheat	1,796,576	1.39	1,765,337	1.55
Soft wheat	611,583	1.08	492,642	1.07
Barley	340,103	1.03	299,653	1.02
Colza	30,589	1.07	22,197	1.07
Maize	1,334,071	1.34	1,382,933	1.33
Sunflower	229,018	1.07	247,850	1.05
Soy been	253,496	1.01	254,740	1.08
Sugar beet	279,981	1.11	245,644	1.16
Tomato	101,229	1.21	87,231	1.13
Tabacco	38,963	1.00	39,263	1.18
Italy	5,015,608	1.26	4,837,490	1.32

6. The use of remote sensing data for optimising an area sample design

The precision of estimates of an area survey is heavily influenced by the sample design. If a previous survey has been carried out, then these data can be used to improve the sample design (Carfagna and Gallego 1995 a, b). One of the ways to use previous data is calculating the correlogram for each observed variable. In fact, an analysis of correlograms gives many suggestions for the optimum segment size and for the number of stages to adopt to maximise the precision of estimates under a fixed budget and a given cost function (Carfagna 1998, Gallego 1998).

In a similar way, when data have not been collected by a previous sample survey, photo-interpretation of remote sensing data can be used to calculate correlograms (see Carfagna, 2000). Attention should be devoted to the fact that, if the correspondence between photo-interpreted classes and the variables of interest is not perfect, then the results can be influenced. Remote sensing data are much cheaper than ground surveys; moreover, since for designing an optimum sampling plan it is not strictly necessary to use very recent data, remote sensing data acquired for other purposes can be used.

7. Conclusions

Remote sensing is very useful for producing agricultural and environmental statistics and can be used in various ways. The most famous and debated use is as auxiliary variable to improve the precision of estimates. Its cost-effectiveness has been debated for a long time, depending on many parameters, such as the level of fragmentation of

the landscape, the weather conditions, the level of optimisation and automation of the project, the cost structure and so on. Thus, different results have been reached in different experiences. Anyway, a drastic change in the pricing policies of Landsat TM data ensures nowadays a cost-effectiveness use of these remote sensed data.
Then, when evaluating the cost-effectiveness of the use of remote sensing data, it must be taken into account that remote sensing provides more than an improvement of the statistical precision, since an information on the location of the crops, or other land uses, is also given. Finally, remote sensing data are also used for area frame construction and stratification and for designing optimum sample plans. For these applications, cost-effectiveness is easily reached.
An important problem that rises, when setting up a project using remote sensing data, is the need of specific skill in GIS, computer science, photo-interpretation, statistics and so on. Indeed, the USDA and the ITA consortium have considerably improved the automation of some parts of the process and some more efforts could be made.

References

Allen J.D. (1990) A Look at the Remote Sensing Applications Program of the National Agricultural Statistics Service. *Jour. of Official Stat*. VI, 4, 393-409.

Brun C., Delincé J., Léo O., Porchier J.C. (1992) Pilot use of the TER-UTI data in agricultural statistics procedure using remote sensing, *The application of remote sensing to agricultural statistics*, Office for Official Publications of the E.C., Luxembourg.

Carfagna E. (1998) Area frame sample designs: a comparison with the MARS project, *Proceedings of Agricultural Statistics 2000*, International Statistical Institute, Voorburg.

Carfagna E. (1999) The use of area frames and remote sensing for agricultural statistics, invited paper at *Meeting on Food and Agricultural Statistics in Europe*, Geneva, 23-25 June 1999. Proceeding on the web site: http://www.unece.org/stats/documents/1999.06.agri.htm

Carfagna E. (2000) Note sull'uso dell'autocorrelazione spaziale nel disegno campionario areale, invited paper at *Disegni campionari per l'analisi spaziale - XL Riunione scientifica SIS 2000*, Firenze, 26-28 April 2000. Proceedings in press. Rome, 169-182.

Carfagna E., Gallego F.J. (1995 a), Sulla dimensione ottimale del segmento per rilievi a terra e fotointerpretazione, *Statistica e Telerilevamento*, AIT Quad. n. 3, CNR, Pisa.

Carfagna E., Gallego F.J. (1995 b), Extrapolating Intracluster Correlation to Optimize the Size of Segments in an Area Frame, *Conference on Applied Statistics in Agriculture*, Manhattan, Kansas, 24-26 April 1994.

Chhikara R.S., Houston A.G., Lundgren J.C. (1986) Crop acreage estimation using a LANDSAT-based estimator as an auxiliary variable, *IEEE transactions on Geoscience and Remote Sensin*g, vol. GE-24 1, 157-168.

Cochran W. (1977) *Sampling Techniques*, Wiley, New York.

Consorzio ITA (1987) *Telerilevamento in Agricoltura, Previsione delle Produzioni di Frumento in Tempo Reale e Sviluppi Tecnologici*, Ministero dell'Agricoltura, Roma.

FAO (1998) *Multiple frame agricultural surveys*, Vol. 2, FAO, Rome.

Gallego F.J. (1994) Using a confusion matrix for area estimation with remote sensing, *Il telerilevamento spaziale per l'ambiente e il territorio in Italia: dalla ricerca al servizi*, Rome, 1-4 May 1994

Gallego F.J. (1998) On the feasibility of diachronic regression estimators with ground survey and Landsat TM data. *International Journal of Remote Sensing*. 19, 8, 1621-1625.

Gallego F.J., Carfagna E., Fuenette I. (1999), Optimizing the size of sampling units in an area frame *geoENV II - Geostatistics for Environmental Applications*, Gomes-Hernandez J., Soares A., Froidevaux R. (Eds.), Kluwer, Dordrecht, 393-404.

Gallego F.J., Carfagna E., Peedell S. (1999) The use of CORINE Land Cover to improve area frame survey estimates, *ROS - Research in Official Statistics*, II, 2, 99-122.

Gallego F.J., Delincé J., Rueda C. (1993) Crop area estimates through remote sensing: stability of the regression correction», *International Journal of Remote Sensing*. 14, 18, 3433-3445.

Giovacchini A., Brunetti A. (1992) Agricultural Statistics by Remote Sensing in Italy: an Ultimate Cost Analysis, *The Application of Remote Sensing to Agricultural Statistics*. Off. Publ. of the E.C. Luxembourg

González F., Cuevas M. (1993) Remote sensing and agricultural statistics: crop area estimation through regression estimators and confusion matrices, *International Journal of Remote Sensing*, 14, 6, 1215-1219.

Hay A.M. (1988) The derivation of global estimates from a confusion matrix, *International Journal of Remote Sensing*, 9, 1395-1398

Hay A.M. (1989) Global estimates from confusion matrices. A reply to Jupp. *IJRS*, 10, 1571-1573.

Jupp D. (1989) The stability of global estimates from confusion matrices. *IJRS*, 10, 1563-1569.

Meyer-Roux J., Vossen P. (1994) The first phase of the MARS Project, 1988-1993: overview, methods and results, *The Mars Project: overview and perspectives*, Office for Publications of the EC. Luxembourg, 33-81.

Taylor J., Sannier C., Delincé J, Gallego F.J. (1997), *Regional Crop Inventories in Europe Assisted by Remote Sensing: 1988-1993, Synthesis Report.* Office for Publications of the EC. Luxembourg.

7 June 2001

Invited Paper Session

STATISTICAL DATA EDITING: METHODS AND SOFTWARE

Chair: G. Hanuschak

Available Methods, Techniques and Software for Survey Data Editing

R. Benedetti, F. Piersimoni
Istat - Servizio Agricoltura, Via Ravà, 150 - Roma, Italia
e-mail: benedett@sci.unich.it

G. Espa
Università di Trento, Via Inama, 5 - Trento, Italia
e-mail: gespa@risc1.gelso.unitn.it

Abstract: The purpose of this paper is to show that a complex sampling structure of the surveys concerning farms, conducted by official statistical institutes in order to gather information on the agricultural sector usually have important implications for the improvement of data editing. The tools most frequently used in this delicate phase of data production are not, as will be shown, sufficiently flexible to adapt adequately to an agricultural statistical system. Consequently, the ISTAT agricultural service has developed a generalized system for panel surveys (AGAIN) whose data editing module has been designed to fully exploit the sample designs developed to better use auxiliary information, and to facilitate the detection of possible errors in the raw data.

Keywords: Statistical Data Editing, Outliers Detection, Graphical Interface, Object Oriented Programming

1. Introduction

As happens in any statistical survey, both of a census and sample nature, and also in the complex system of surveys relating to the Italian primary sector, the problem of error detection and correction in the obtained data has to be considered. To act automatically on the errors and/or to rebuild the missing information, will save time and computational costs needed in the data management, and will increase the quality of disseminated information. This last item assumes particular importance if we consider that in the surveys relating to the farm activity, the units where errors are discovered are surely not a sub-sample of the farms interviewed. In most of the applications, errors are not random. For instance, farms of large dimensions will give more importance than small farms to the utility of the survey, having a greater impact on the accuracy of the data produced. In addition, as the economic data often present very relevant asymmetries, we realize that also a small quantity of errors can have a large impact on the estimates, if such quantity is due to the most important farms. In this case, therefore, it is essential to make sure that errors are identified and corrected in the most accurate way. To the above requirement has to be added the review that Istat (Italian National Institute of Statistics) has set out for its farms surveys. The need to overcome an information plurality, unfortunately persistent in Italy, has compelled Istat to study an

articulated and complex survey structure based on the logic of sub-samples selected from a "master" survey used as reference frame. In this phase of redesign of the existing surveys, a data control and correction strategy, taking into account changes introduced, has to be prepared. The problem to choose between two alternatives has then become apparent: to develop *ad hoc* systems for the new situations created, or alternatively to utilize techniques already in existence. The purpose of this work is therefore to make a comparison in quality (rather than quantity) terms of the functionality of some of the most diffused data check and correction systems. Such a comparison will allow us to confirm and justify the Istat agriculture service choice to develop a product alternative to those existing, mainly oriented to *panel* surveys. It deals with the data check and correction module contained in the *software* AGAIN (*Analisi e Gestione Automatica delle Indagini*), realised for a global management of several surveys.
The structure of this work is as follows. Paragraph 2. will briefly introduce the new system of the Istat primary sector surveys that is being defined after the revision above quoted. Section 3. will contain a critical presentation of some of the most diffused data control and correction systems. This paragraph is subdivided into two subsections: the first concerns GEIS. The second subsection is dedicated, with fewer details, to the review of other important products. Paragraph 4. will describe the main distinctive features and functionalities of the system developed by the agriculture service of Istat, with particular reference to the logic (different from other products) of data analysis for outliers identification. Section 5. will contain some conclusive comments.

2. A survey system for data editing

Farms surveys have been based, for years, on frames, methods and techniques not coherent among them (Espa *et al.*, 2000). Above situation has prevented for long time to satisfy the minimum requirements of integration or at least of comparison of the obtained estimates. Such considerations led Istat to study a complex structure of surveys based on the logic of subsamples selected from a "master" survey taken as reference frame. The survey on the structure and production of farms that Istat is now performing yearly, seems naturally addressed to implement such function. In fact, many of the questions included should be in any case inserted in any questionnaire because they are aimed to know general aspects of the management type and of resources utilized for production (surfaces dedicated to the various cultivations, heads of cattle, personnel and production means utilized). Moreover, this survey has such dimensions as to make it possible to utilize the above subsamples to produce estimates on variables that are not those for which the structure survey has been designed.
To decrease organizational and managemental costs, Istat has introduced longitudinality in the structure of the "master" survey, utilizing a panel with yearly rotation for each stratum of 20% of the units. The hypothesis behind this choice is that maintaining about 80% of the units in the sample from one year to the other, also if it complicates the sample structure, allows easier identification and localization of the farms with consequent reduction of the survey execution time.
From a methodological point of view, it has to be pointed out that the above sampling design, although apparently complicated, provides some significant advantages. The difficulties that come from the development of the procedures of computation of the

sample sizes and of random selection, have their compensation, as can better be seen in paragraph 4., in a series of possible developments in the data editing, that have a significant impact on the quality of the estimates. Before proceeding in that direction, however, it is necessary to examine the functionality of the most diffuse data control and correction systems, in order to show their inadequacy for their application to the described survey design. In fact, such structure must not be considered as a "stand alone" entity, but rather as a methodological investment, addressed to the reduction of the number of sampling errors and to the dissemination of new databases built thanks to the integration of several data sources (Espa *et al.*, 2000).

3. Systems for the detection and correction of errors

3.1 GEIS-Generalized Edit and Imputation System

The GEIS system has been developed by *Statistics Canada* in the framework of the *Generalized Survey Function Development Project* in order to answer requirements and problems arisen in the planning of the Canadian economic surveys.

GEIS uses a control and correction procedure for quantitative variables, based on the theory of Fellegi and Holt (1976) and, in order to solve optimisation problems defined by variables collected in statistical survey, uses a generalization of the Chernikova algorithm (1965).

The process of identification of the fields to impute and of automatic correction is guided by a set of linear rules defined on quantitative variables. More details on GEIS structure and functionality are contained in Statistics Canada (1998). GEIS utilization implies in any case that part of the follow-up and/or control activities, such as manual data correction, have been made in a preliminary phase of editing (especially in relation to systematic errors).

In a synthetic manner, The Fellegi-Holt methodology included in GEIS can be outlined by figure 1. Of course, if after the application of the described procedure, some record still fails some or more rules, the steps (2)-(4) can be repeated, the errors localization phase being limited to the consideration of the only variables already imputed. Step (1) allows the identification of axioms defining the acceptable records. Such rules (edits) are expressed as n linear constraints as follows:

$$\sum_{j=1}^{m} a_{ij} x_j \le b_i \text{ or } \sum_{j=1}^{m} a_{ij} x_j = b_i, i = 1,2,...,n \qquad [1]$$

(in which positive constraints $x_j \ge 0, \forall j$ have to be added) where constants a_{ij} e b_i are specified by the user while x_j are the *m* sample variables. The system [1] is therefore made up by a set of *n* rules connected by the logic operator "and", and this means that a record has to satisfy each single edit before passing the control phase. In the implementation of the editing strategy, an edit analysis phase has to follow that of specification (cfr. step (1) of Fig.1). Referring to the above phase, we intend to outline the control made on the edit, in order to build the "minimal consistent" set of edit, which will be utilized in the error identification phase. The control step consists in the search of possible

1) inconsistencies: the acceptance region delimitated by edit may not be empty. The elimination of inconsistencies is made by attempts;
2) redundancies: an edit is redundant if it doesn't reduce the acceptance region. The redundant edit is directly eliminated;
3) hidden equalities: it is a question of equalities implicitly contained in the set of the edit.

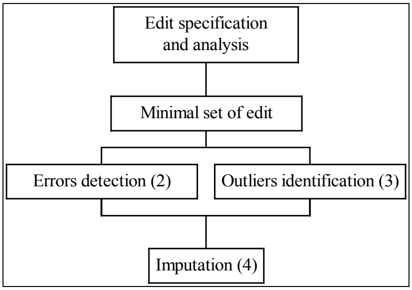


Figure 1: *Steps composing the Fellegi-Holt methodology included in GEIS.*

The second step consists in the application of a particular control function in order to identify the univariate outlier. It is the procedure proposed by Hidiroglou and Berthelot (1986) that identifies the outlier relative to a variable, identifying those values exceeding the prefixed acceptance interval.

The third step is the error detection, i.e. the identification of those records activating one or more edit. For each such record the fields to be corrected are identified in such a way that:

1) the number of the necessary changes is minimal;
2) the modified record verifies all edits.

In GEIS such problem is expressed as a linear programming problem, with the constraint of a solution with minimal cardinality. The above problem, as already said, is solved utilizing the Chernikova algorithm (1965) in the Schiopu-Kratina and Kovar formulation (1989). In the GEIS system is also possible to use weights to be allocated to some variables to which more importance, for many reasons, has to be given.

As far as the final step is concerned, GEIS foresees three imputation strategies:

1) *deterministic*. Inside each record, for each variable to be imputed, it is verified if there exists one and only one value that, once fixed, makes the record to satisfy all the edits;
2) *with donor*. For each wrong record the "nearest" record donor is identified, in relation to the minimax distance, from the so-called coupled variables to be chosen in advance;
3) *with estimators*. The variables identified during the error identification phase, are imputed one at a time utilizing various types of estimators (ratio, current or previous average, trend).

The advantages offered by GEIS utilization are numerous, justifying its popularity and large diffusion (it is almost a standard) in the specific work environment. Just to quote the most relevant qualities, it is enough to remember that GEIS is a system:
1) modulable and therefore it is possible to utilize again parts of it in other work situations;
2) flexible, especially as far as the estimation module is concerned;
3) that allows an efficient monitoring of the record situation, in particular through production of a series of diagnostic reports, relating to the file under management.

On the other hand it has to be said that GEIS:
1) is not at all user-friendly because it requires that the user be familiar with Oracle environment. Moreover, up to now, utilization of SAS® is not allowed;
2) doesn't allow direct access to an imputation module that a user needs, as happens very often, to utilize a system of external edit;
3) foresees that each variable subject to control and correction assumes non-negative values. This very often requires some preliminary adaptations;
4) doesn't produce graphical displays of the data being monitored;
5) offers, according to us, a too limited degree of interactivity.

In conclusion, the major advantage offered by GEIS utilization is that such system is generalized, i.e. it is applicable to a large range of economic-administrative surveys. On the other hand, this quality becomes a big weakness when we leave the structural surveys area and there is the need to work on surveys, carried out several times on the same units. In fact GEIS is not fit for being adopted in a suitable way to the panel optics often used in conjunctural surveys. For that reason there was at Istat Agriculture Service the need to develop an ad hoc *panel-oriented* product (the data editing module contained in AGAIN), that allows the creation of a relation with the new surveys system planned for the primary sector.

3.2 Other products

NIM (New Imputation Methodology)

There is another system developed by Statistics Canada, addressed to the social area surveys and oriented, therefore, mainly to the treatment of qualitative variables. It uses only one imputation method based on a donor, identified with a function of mixed minimal distance. Details on this subject can be found in Abbate (1997). NIM is more an imputation than an editing system and therefore it is used after the set-up of edit rules. For this purpose it is possible to use SPIDER, a package also produced by Statistics Canada, that can be interfaced with NIM.

NIM has been used with some success for the correction of data in the Canadian population Census in 1996 (Bankier *et al*., 1997), but certainly implies a consistent quantity of generalization work for further applications. It is, in any case, a product not sufficiently tested, especially for the situation of performance evaluations, with an increase of quantitative variables inside the data set to be analysed (Statistics Canada, 1999). NIM has mainly two targets: to correct the minimum number of fields and at the same time to suggest an acceptable imputation.

StEPS (Standard Economic System Processing System)

StEPS has been projected with the ambitious aim to substitute 15 pre-existent systems for the American economic surveys (U.S. Bureau of the Census, 1996). It is much more than a data control and correction system because it provides, as set-up for AGAIN as well (cfr. paragraph 4.), integrated tools for analysis in various phases of a survey. The editing module of StEPS allows both very simple checks (presence of specific values, ranges, category validity), and more complicated tests. Moreover, the user can submit in SAS® (StEPS is entirely developed in SAS®) its own program that imposes control rules on current and historical values. The SAS® utilization privileges statistics users that are expert in surveys, because they can exploit the enormous potentialities of such software. StEPS moreover allows an interactive manual data modification, that can be repeatedly run by more users, when data are not submitted to automatic imputation. The imputation module is made up by two sub-modules, the first one being a classic module of deterministic imputation (simple imputation). The second one, in contrast, is a module of general user-defined rather than automatic imputation, as for GEIS and NIM. The imputation techniques available in such module are substantially based on the utilization of estimators computed from weighted or unweighted data. Therefore in the imputation strategy it is not possible to impose the rule of minimum change foreseen in the Fellegi-Holt methodology, as well as using an automatic function based on some donor criterion. Finally StEPS runs in a Unix environment and is provided with a graphical interface.

Solas
Solas is a product by Statistical Solutions Ltd. in Ireland. Initially, it has been developed for imputation of missing data in the biostatistic field. Even if it is oriented to the treatment both of qualitative and quantitative data, Solas reaches a better performance on quantitative data. It should be noted that such a system can work both on sectional and longitudinal data. Solas doesn't include editing functions as it only imputes missing data. In this respect it has a very sophisticated module with multiple imputation (see Rubin, 1996 for a detailed explanation), and it foresees the error correction by means of the standard hot-deck random method and two types of imputation via an estimator (current average – mode for qualitative variables – and historical imputation). Solas runs on a PC 486 and higher, and can read and write data in various formats. The Windows interface is user-friendly and it is very similar to the SAS/INSIGHT® module. On the other hand Solas doesn't produce a synthesis report on the correction process. In conclusion it is, in our opinion, a product not suitable for complex surveys, in which a high quantity of variables linked by complex internal relations are involved.

AGGIES (AGricultural Generalized Imputation and Edit System)
AGGIES has been developed by the U.S. *National Agricultural Statistics Service* (NASS) with the aim to build an efficient and objective automatic system to control and correct agriculture data. Before developing AGGIES, NASS has performed accurate research in order to identify the best product to be introduced into the agricultural environment for editing and automatic imputation, based on the Fellegi and Holt (1976) model. After evaluation of some alternative solutions, it has been decided that the GEIS paradigm was the most interesting, especially because it was a generalized system. The subsequent step was to develop a product with the GEIS characteristics, but written in SAS® (in the specific SAS/AF® and SAS/IML®), taking into account the experience gained during years by NASS with this language.

AGGIES structure is modular and each component module is dedicated to a specific function. The AGGIES main functions are the following:

1) edit specification: these are expressed in a linear form (cfr. formula [1]);
2) identification of data and/or edit groups: AGGIES includes a module for application of a set of edits (i.e. a selection of the complete set) to a particular data group (i.e. to data of a stratum of the sampling design or of a region);
3) edit verification: see the control step described for GEIS (par. 3.1) and/or Giles (1988);
4) edit synthesis: this is a module that uses data to be controlled and corrected. The output of this module shows, for each edit, the number of records that pass the set of edit rules and the number of those that fail at least one rule;
5) outliers identification: this is a module similar to the correspondent developed in GEIS, as far as the used methodology is concerned;
6) error identification: a module similar to that implemented in GEIS (cfr. par. 3.1). Even in AGGIES it is possible to give weights to specific variables. As in GEIS an high weight given to a variable signifies a high confidence in the values that the variable assumes and therefore a low probability that those values will change;
7) imputation: AGGIES provides two imputation strategies: by donor and with estimator. If no value imputed with an estimator can satisfy all the edits, AGGIES outlines the set of values of the variable concerned that satisfy all the rules and then automatically imputes the average of such values. The output tables which follow the imputation phase are standard and can help, together to those produced during phase 5), during the subsequent phase of revision.

Advantages and disadvantages are almost the same as those already described for GEIS, because AGGIES is substantially a SAS® version of GEIS itself. However, it does not satisfy the need, much felt in NASS, of a more consistent integration of AGGIES with other tools for editing and analysis already tested by NASS itself (SPS for micro-editing, Blaise for interactive editing and IDAS for macro-editing), in view of the preparation of a complete strategy of control and correction on the surveys relating to the U.S. primary sector.

4. The solution of ISTAT's agricultural service: the AGAIN system

AGAIN is a generalized tool for the complete and unitary management of ISTAT surveys on agriculture. This section will concentrate almost entirely on AGAIN's data editing module, while specific details on all the modules of which AGAIN is composed are provided by Benedetti and Piersimoni (2000). AGAIN allows the inclusion in the system of methodologies other than those implemented, and thus, in modular manner, enables the user to do the following:

1) design the sample. The procedure is multivariate and uses, in the case of stratified sampling, a generalization proposed by Bethel (1989) of Neyman's classic formulas for the calculation of optimum numerosity;
2) produce estimates and relative sampling errors. This module enables estimation of the totals of each variable of interest, as well as the relative sampling errors, by means of the Horvitz-Thompson estimator. It also permits adjustment of the basic sampling design weights by total non-response, producing the so-called 'base

weights'. These weights can be adjusted further in AGAIN by means of calibration (Deville and Särndal, 1992), so that the current survey estimates are coherent both with those produced by other surveys and with the information yielded by the reference frame;

3) data editing. This module will be described in more detail in what follows.

Before the data editing screen is described, a specification is necessary. Perhaps the most significant problem that arises during the phase of detecting and correcting the information collected by a survey is not so much the particular imputation system used as the strategy selected to detect errors. In this regard, the AGAIN module is based on a data 'navigation' approach, rather than on an algorithmic structure. It was decided not to use a Fellegi-Holt approach, for two main reasons: (i) when the questionnaire is very complex, a large number of initial (explicit) edits may be detected and, more importantly (ii) only if the errors present in the data-set to be checked are stochastic in nature does this method detect errors in an optimal manner.

Once the errors have been detected, the fields concerned are transformed into missing data. Only now can the more or less structured phase of reconstructing the missing data begin. Since partially missing data are practically non-existent in the surveys considered here, the problem of the missing fields to be estimated is generated following error detection. As shown more clearly below, this aspect may constitute an even more complex phase of analysis. However, the specific nature of the variables used in surveys of farm means that the formal rules underlying the data during the error detection phase are not excessively numerous. Indeed, because of the manner in which the phenomena are structured, this preliminary editing phase concerns not so much identification of the circumstances in which rules are breached as identification of possible outliers. Identification of the latter necessarily requires the use of external sources. This auxiliary information, as we will see, is available thanks to the complex and structured design of the survey system, whose longitudinal structure permits the use of data from previous years, or from other surveys, in order to conduct more precise checks on each single field. Turning to a more detailed description of the system, AGAIN has a single data detection and adjustment panel consisting of seven linked sub-screens, which can be opened simultaneously. The first five of these sub-screens allow the user to 'navigate' through the data and select them interactively, while the last two are the correction modules. The operations are extremely easy to perform through AGAIN's implementation of SAS® procedures which, by virtue of the new technological frame introduced in the SAS/AF® module, ensure high interactivity and high-resolution graphic images.

The first panel ('Scatter') enables detection of suspicious values by means of simple visual examination of the scatter of points between the variable being checked and the same variable relative to a past period (see Fig. 2a).

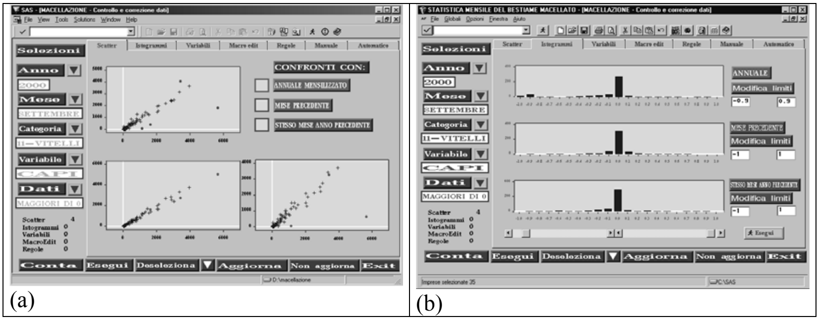
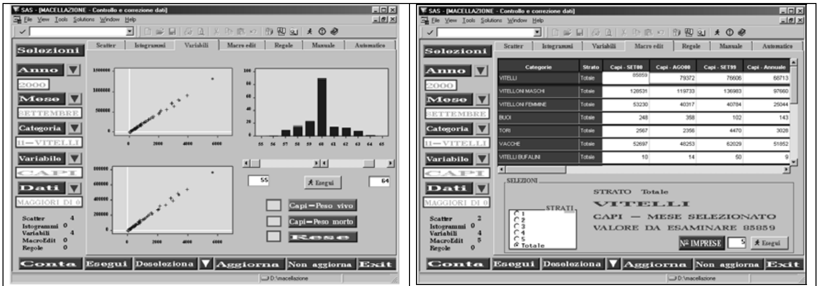


Figure 2: *(a) outliers identification by means of temporal scatters, and (b) second screen: identification of the outliers by means of histograms.*

The three scatters refer to three different cases of phase displacement against which the values of the current variable selected can be compared. This comparison can be made against annual values, against those for the previous month, or against those for the same month of the previous year.
Obviously, before the graphic representations are constructed, the information to be checked is selected from the reference frame by means of the 'Selezioni' sub-screen on the left side of the monitor. This selection can be made relatively to a reference month or year, or by a combination of items contained in the 'Categoria' and 'Variabile' sections. Moreover, it is possible to show all the data present in the vector or only those with positive values.
By means of a simple exploratory analysis of the data guided by the scatters, the operator now selects the suspicious values by clicking on those points of the scatter which he judges to be 'out of trend'. The records corresponding to the points selected can be displayed in the sixth screen ('Manuale'), which is used for the manual correction of errors (see Fig. 4b). Subsequent selections are added in the 'Manuale' screen to those previously made and can be excluded only if deselected. The deselection operation is performed in two ways: (i) directly on the scatter by clicking again on a point selected so that the initial 'normal value' state is restored or (ii) by using the 'Deseleziona' command located on the horizontal tool bar at the bottom of the panel: this option can be used to deselect all the records identified, or else only those selected in a particular data detection screen.
AGAIN's second data editing screen displays the histograms of the normalized variations in the interval [-1, 1] (see Fig. 2b). These histograms are relative to variation with respect to a reference period t_0, which can be selected from the alternatives shown in the first screen. By selecting one histogram at a time, and by altering its limits simply by moving the horizontal slide bars, the operator isolates the elements on the distribution codes. This operation identifies a number of units with 'suspicious' values for the variable in question, and if they are not deselected, appear automatically in the manual correction screen. The first two screens should be used jointly, because they have different capabilities and information content. In fact, by means of exploratory analysis of the scatters it is possible to detect the 'large outliers', while the opposite

happens if one concentrates on examination of the histograms (location of the 'small' outliers).
The third sub-screen ('Variabili') contains the same kinds of graphic representations as the previous screens, with the difference that these graphics do not involve intertemporal comparisons but are used to cross-reference – only for the current period – the variables for each category (see Fig. 3a).

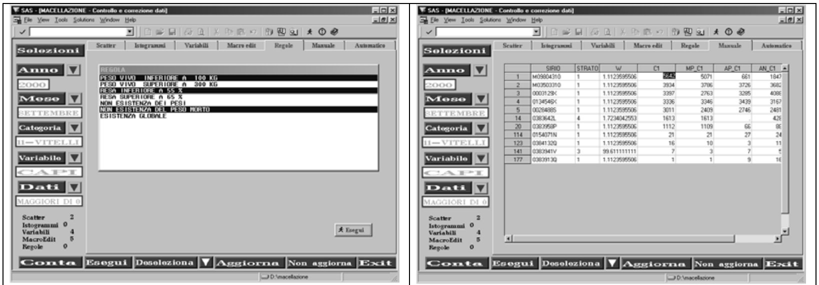


(a) (b)

Figure 3: *(a) third screen: relationships among different variables relative to the same period; (b) fourth screen: macro-edit.*

As we know, quantitative variables express magnitudes which assign different weights to the statistical units in a population. For example, the distribution of the values of the variable *number of heads* for a certain category of livestock, say *calves*, shows a concentration of the total quantity in a given fraction of farms. As regards the errors detection of which variables of this type may be affected, it is obvious that those located in the 'more important' farms have a greater impact on the total for the reference population. It is therefore necessary for these errors to be detected and corrected as accurately as possible. It is for this reason that AGAIN includes a macro-editing screen (see Fig. 3b) based on the following principle: the manual follow-up and control activities are concentrated on the farms with the greatest impact on the final estimates by operating on the others by means of automatic procedures. In other words, the following method is used. First, the value of the 'suspicious' aggregate to be examined is selected from the spreadsheet-type screen. This selection may concern the estimate of the total of the variable of interest or the same type of estimate produced for one of the layers into which the population surveyed has been divided. Moreover, the selection may be intended to isolate the records with the greatest impact on the absolute aggregates or on their percentage variations. At this point, the individual data with their final weights are arranged in increasing order. The 'tail' of the distribution thus produced is now selected (i.e. the user specifies the number of units to be selected), and the data with the greatest impact on the aggregate considered are examined manually (in the 'Manuale' screen) (Granquist, 1990).
Considering the time and the resources available for surveys in the agricultural sector, the procedure just described cannot be applied to all units. It is therefore followed by automatic correction techniques on the less relevant units influencing the final estimates.

The next screen ('Regole') displays the edit rules that apply to the variable under analysis (see Fig. 4a). The user may decide whether to impose only some of these rules, or whether to apply the complete set of edits.



(a) (b)

Figure 4: *(a) fifth screen: the rules; (b) sixth screen: manual correction.*

The sixth screen ('Manuale') consists of an editable table which displays the 'suspect' records selected for manual correction in one of the previous screens (see Fig. 4b).
The seventh and final screen ('Automatic') is the most composite one (see Fig. 5), it provides algorithms for the imputation of one variable at a time, with the possibility of using different estimators. The correction methods available are listed in the upper part of the screen, while the parameters to be specified for each method are grouped in its lower part.
It should be noticed that, since the estimators are applied independently for each variable, the records 'cleaned' on conclusion of the automatic correction phase may not satisfy the compatibility plan. It will therefore be necessary to perform the error detection phase again using the tools described above.

Figure 5: *seventh screen: automatic correction.*

The estimators implemented in AGAIN are the standard tools usually used for automatic imputation: previous value and average, current value and ratio, auxiliary trend and difference, regression, donor. For the last two estimators, it is also possible to act on more than one temporal delay. In the former case (regression estimator), the number of regressors increases, while in the latter the distance between 'recipient' and 'donor' is calculated with respect to certain items of information about all the delayed values.

5. Conclusions

The advantage of AGAIN with respect to other data editing systems, is that it has a modular and interactive structure for data exploration which detects suspicious values and allows their manual or automatic correction. The detection phase essentially consists of an elaborate process of exploratory data analysis, and is made easier through high-resolution graphic displays, the execution of which does not require statistical expertise by the operator. The subsequent adjustment phase is relatively flexible and allows for various strategies of intervention (manual or automatic) on data with respectively greater or lesser relevance to the final estimates. To conclude, it seems that the great effort made to harmonize the agricultural surveys conducted by ISTAT yielded important results with AGAIN. In fact, during the data editing phase, AGAIN fully exploits all the integrated information available, giving flexibility and reproducibility to an estimation process which thus becomes less subjective and more controllable.

References

Abbate C. (1997) La completezza delle informazioni e l'imputazione da donatore con distanza mista minima, *Quaderni di Ricerca Istat*, 4, 67-102.

Bankier M., Houle A.-M., Luc M., Nadeau C., Newcombe P. (1997) 1996 Canadian Census Demographic Variables Imputation, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Benedetti R., Piersimoni F. (2000) Nuove tecnologie per laproduzione e diffusione dell'informazione statistica sui fenomeni economici, nota presentata al covegno *Captor 2000 – Sistemi statistico-informatici per valutare la qualità della didattica*, Padova, 14-16 dicembre.

Bethel J. (1989) Sample allocation in multivariate surveys, *Survey Methodology*, 15, 47-57

Chernikova N.V. (1965) Algorithm for finding a general formula for the non-negative solutions of a system of linear equations, *USSR Computational Mathematics and Mathematical Physics*, 5, 228-233.

Deville J.C., Särndal C.E. (1992) Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, 87, 376-382.

Espa G., Benedetti R., Russo M. (2000) Alternative metodologiche per la stima di statistiche agricole provinciali, Atti del convegno SIS *Verso i censimenti del 2000*, Udine, 7-9 giugno 1999, II, 335-345.

Fellegi I.P., Holt, D. (1976) A systematic approach to automatic edit and imputation, *Journal of the American Statistical Association*, 71, 353, 17-35.

Giles P. (1988) A model for generalized edit and imputation of survey data, *The Canadian Journal of Statistics*, 16, 57-73.

Granquist L. (1990) Macroediting – The aggregate method, Statistics Sweden, SCP/2D.13/f, october.

Hidiroglou M.A., Berthelot, J.-M. (1986) Statistical editing and imputation for periodic business survey, *Survey Methodology*, 12, 73-83.
Rubin D.B. (1996) Multiple imputation after 18+ years, *Journal of the American Statistical Association*, 91, 473-489.
Schiopu-Kratina I., Kovar J. (1989) Use of Chernikova's algorithm in the Generalized Edit and Imputation System, Statistics Canada, Methodology Branch Working Paper No. BSMD-89001E.
Statistics Canada (1998) Functional description of the Generalized Edit and Imputation System, *Statistics Canada Technical Report*.
Statistics Canada (1999) A functional evaluation of edit and imputation tools, Working Paper No. 12, *Conference of European Statisticians*, Rome, 2-4 June.
U.S. Bureau of the Census (1996) StEPS: Concepts and Overview, *Technical Report from the U.S. Bureau of the Census*.

Outliers, Inliers, and Just Plain Liars -- Interactive Graphical EDA + (EDA Plus) Techniques for Analyzing/Editing Data

David DesJardins
Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233-9100 USA, e-mail: david.l.desjardins@ccmail.census.gov

Abstract: This paper describes a number of special graphical data analysis/editing techniques that have devised by the author. Graphs, the natural language of mankind, offer even novice data analysts a quantum leap in their data editing/analysis capability. This quantum leap in capability is urgently needed because of our dynamic, rapidly changing world -- where we are faced with a virtual flood of data -- often from very vital/complex systems. Unfortunately, this flood of data can likewise be a real challenge to understand/analyze/edit. For instance, many of our traditional/conventional data analysis/editing methods generate fixed-formula printouts requiring an analyst to review and correct the fields of records that are thought to be erroneous. There are many limitations to these conventional methods -- even when well designed. They sometimes overlook basic methodological problems, they typically channel the reviewers in a manner that either may not allow a number of the errors to be found, or they focus on traditional (outdated?) relationships, and they often do not account for changes in the data.
A special course has been designed by the author to teach a number of powerful interactive graphical data analysis/editing techniques to deal with these problems. This now, very popular, course uses new, easy to learn (point-and-click), interactive Exploratory Data Analysis (EDA) software packages (SAS's JMP and Insight graphical data analysis software) -- and makes these techniques very straightforward to apply. These graphical methods can first be applied in an interactive exploratory manner -- to discover nuances that conventional methods are likely to miss. Then, specially designed graphic forms can then be used in a straightforward, highly productive manner to interactively edit these data. (NOTE: The "plus", "+", is used in EDA+ is used to signify the enhanced interactive EDA methodology developed by the author to enhance EDA methodology -- this is also called EDA PLUS). Lastly, these new EDA+ graphical methods allow "inlier" detection and confirmatory review of data -- to not only find hidden relationships within these data but to also assure that corrections made during to the editing process have worked well.[1]

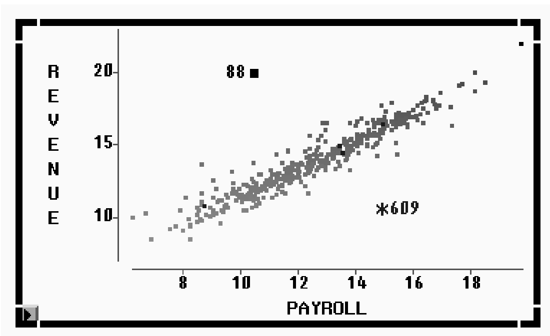
Keywords: Exploratory Data Analysis (EDA), EDA + (EDA plus), interactive graphics, live/dead graphs, inliers; industry profile, data profile, and leverage plots

1 Although these interactive EDA + techniques are also taught as part of a graduate level statistics course by the author, they offer even novice subject matter specialists a quantum leap in their data analysis capability. As such, it has been taught (with considerable success) to students with no statistical background.

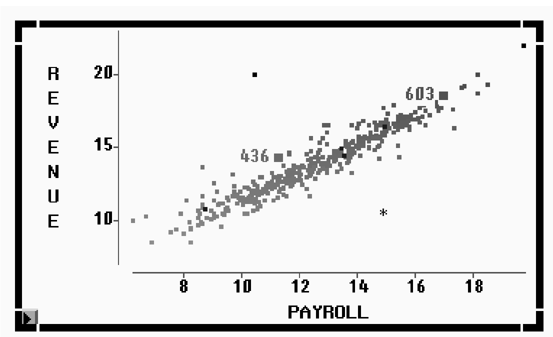
1. Introduction

A key focus of this paper is on ***"Live" graphs*** -- versus ***"Dead" graphs*** -- may they rest in peace!! Graphs, the natural language of mankind, often make even complex statistical concepts easy to understand. Further, even "dead" graphs offer analysts a very powerful data editing methodology. Unfortunately, the vast majority of statisticians and data analysts have only seen/used "dead" graphs to edit their data. This is "unfortunate" because "live" graphs offer analysts an additional quantum leap in analytical capability -- and, with today's relatively cheap, very powerful computer hardware and EDA software, this interactive graphic EDA + (EDA plus) methodology should be available to all. In addition, these interactive methods are virtually ignored by most EDA texts available today.
Four key factors contribute to a revolution in data analysis and make the introduction of these EDA+ methods (with their ***"live" graphs***) at this time a momentous opportunity. First, these graphics software packages provide our analysts with the ability to generate hundreds of "live" graphs in a matter of mere minutes. Generating these graphs would have taken weeks just a few years ago. Second, by using key EDA+ techniques (like brushing and animation) in combination with specially designed graphic forms (like Industry Profile Plots), these ***interactive*** software packages provide powerful/sophisticated "live" graphical methods of looking at the data and reviewing subcomponents of it. If the analysts believe the data is in error, then they can easily discover/correct it in an interactive manner using these point-and-click tools. Third, some versions of this software are incredibly inexpensive. For instance, the comprehensive student version of SAS's JMP, PC-based, software package is available for only about $60. Fourth, and most important, individuals using these multi-purpose EDA techniques are not locked into custom designed software packages and fixed ways of looking at the data. (Many custom designed graphical data analysis packages have a 6-month lead/design window.) By using the above hardware and software tools, we have developed new general-purpose graphical forms and special techniques that greatly enhance the speed and efficiency of data editing/analysis tasks (see particularly DesJardins, 1998). Analysts no longer need spend time editing their data with fixed methods and cumbersome, boring, tabular printouts.
Cleveland (1993) has provided a variety of methods for graphical data analysis. Granquist (1997) and Hogan (1995) have shown how to apply some of the methods to Census data and businesses. This paper shows how a series of well-designed interactive graphical methods (using "live" graphs) have been added to this work -- and are now used to quickly find errors and anomalies in our data. Most graphical and non-graphical methods allow detection of outliers in distributions that may be in error. However, graphical methods can yield insight into situations in which more subtle errors occur. Even the best computer algorithms are "blind" to these errors. Perhaps, the most important point is that these new graphical packages have proven to be very easy to learn/use to locate and correct framing errors and outliers in data -- the basics of the point and click, 4th generation software is typically mastered in just one day.

An example of a data-editing problem that is sometimes erroneously flagged by conventional methodology as an outlier. An *outlier* is a data value that lies in the tail of the statistical distribution of a set of data values. The intuition is that outliers in the distribution of uncorrected (raw) data are more likely to be incorrect. A multivariate analysis (of a few key survey variables) is often required to frame unusually high/low survey responses. Examples are problem data values that are typically flagged by conventional methods are those that lie in the tails of the distributions of ratios of two fields (ratio edits), weighted sums of fields (linear inequality edits), and Mahalanobis distributions (multivariate normal). If graphs are used -- outlying points in multivariate point clouds are typically flagged graphs. However, sometimes only a few variables are used in conventional edits to frame these responses as "outliers". In contrast, our EDA methodology looks at all of the survey responses in an effort to provide our analysts with a real understanding of the nuances of the data.
Below, we see a number of companies that have reported their Revenues and their Payrolls (plotted in log values). In this example, points # 88 and 609 would be considered "outliers".



Points # 436 and 603 correlate well with the majority of these points and would be considered "inliers".



An *inlier,* as opposed to an *outlier,* is traditionally defined as an erroneous data value that lies in the interior of a statistical distribution. Because inliers are difficult to distinguish from good data values they are often very difficult to find and correct using methodology such as the conventional fixed-ratio method mentioned above. A simple example of an inlier might be a value in a record reported in the wrong unit, say degrees Fahrenheit instead of degrees Celsius. A more complex example would be companies who invent reasonable responses to the questions on our survey forms (***"liars"***). Isolated inliers may not be a problem and may be almost impossible to distinguish from correct data. Sets of inliers of moderate size may seriously affect uses of microdata. In some situations, sets of inliers may arise when two or more administrative lists are linked and some of the identifying information is in error. In some situations, we can use graphical EDA methods to discover these sets of inliers. In addition, if corrections (such as imputed data) are done, then we can also use these graphical EDA methods to confirm the plausibility of the changes.

In this paper, we provide an overview of these graphical EDA methods. We show how they can be used to clean up data or discover if there are methodological errors in it. In the next section, we cover how EDA methods can be used to detect simpler errors. Although these methods are often quite straightforward to learn and apply, they often yield information about serious errors that have previously gone undetected (using our conventional methodology). In the third section, we show how EDA methods can indicate the existence of more complicated problems. For instance, these inliers/problems might arise as a mixture of two distributions or of two very different types of companies that appear to be the same on the surface.

2. Basic data/outlier editing using EDA graphs

Conventional editing methods often involve the development of if-then-else rules in computer software -- rules that delineate records that may need editing. These methods have typically been used in computer environments in which printouts are created for analyst review. Review and correction of these data can be time-consuming. Errors may not be located because edit rules are inflexible or are not designed to detect certain classes of mistakes. Updated sets of questions may be asked for a survey. Methods that are developed for one survey may not be exactly appropriate on other surveys. Worse, "error" criteria for our data may change as industries change in our dynamic society.

The use of modern, graphical data software in an exploratory manner allows analysts to detect errors that cannot be detected by inflexible editing rules. These software packages also provide easy methods for updating databases as errors are corrected. In spite of these advantages, many Statistical agencies do not widely utilize these methods. Sometimes their analysts do not know how easily these new methods can be applied -- or, sometimes, introducing change to a bureaucratic agency is just plain hard to do. Sometimes, they may simply not be aware of new, specially-designed data editing training courses -- such as the ongoing EDA courses currently being taught by the author at the U.S. Census Bureau.

Fortunately, these courses have now been taught by the author in such diverse forums as Statistics, Canada, Statistics Italy, the United Nations, and to other statistical groups in a short course format. (These courses use the SAS Institute's powerful, highly interactive, easy to learn (point-and-click) Insight and JMP software.)

EDA methods often allow analysts to discover other "hidden" errors in data -- such as inliers. Fixed computer algorithms (with conventional edit methods) often miss unanticipated errors in the data. The fixed algorithms are blind in the sense that they cannot adapt to new situations. For instance, comparing variables and using ratios requires a very good understanding of the often unique/unpredictable relationships between each of the variables. These relationships can vary markedly for different point cloud clusters associated/corresponding to companies in different industrial codes. These variations can be substantial between companies in different size ranges or when survey forms are re-designed. Time series variances such as business cycles and periodic anomalies can also affect these relationships. Basic relationships for data sets can also vary substantially over time as well. In addition to producing outliers, these factors often hide inliers. By using powerful graphics tools in combination with a subject matter specialists trained in EDA methodology, these problems can be quickly identified and dealt with.

In the next section, we show how these EDA methods have provided us with a quantum leap in our data analysis capabilities. This section highlights key examples of where EDA has shown itself to be significantly better at data editing in the analysis of data, and for identifying outliers and inliers. In addition, we show an example of how EDA is much better at discovering undetected flaws in our traditional data analysis methodologies.

2.1 Example 1 – Basic Day-To-Day Data Editing -- Control Charts

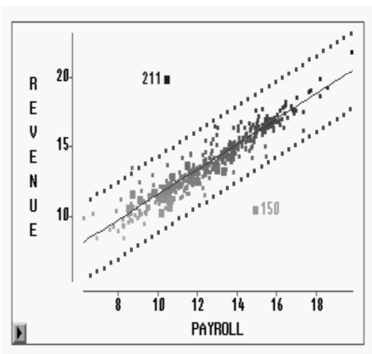
Graphs communicate across a wide area of expertise. Again, a properly chosen graph can often make sophisticated statistical concepts clear to laymen. Thus, using graphs, Statisticians can now more quickly/effectively explain to our subject matter specialists the fundamental concepts behind their data analysis techniques. For instance, the conceptual framework of the graphs shown below (that are now automatically generated by our data editing software), are easy for even a layman to understand. (NOTE: For the purposes of realism, the graphs in this paper are direct copies of the output of our SAS Insight and SAS JMP software packages.) Again, these techniques can be very easily implemented at minimal cost. SAS Insight is available as an addition to the basic SAS software package. SAS JMP is available as standalone software for Macintosh and Windows PCs. A very comprehensive, student version of the SAS JMP software (with a 500+ page statistical data analysis manual) is available for only about $60. The commercial (almost identical) version of this package typically costs about $1,000!

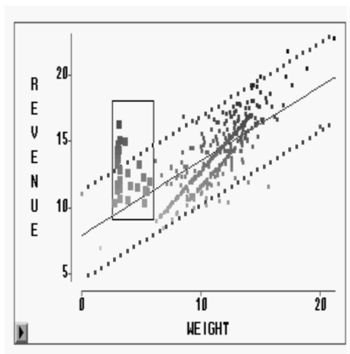
NOTE: For confidentiality reasons, all the examples used in this paper contain artificial data -- which, for illustration purposes, closely mimic our actual data.

The first rule of EDA is to simply plot the data (and then really look at the graphs). The second rule is to note/understand the basic relationships (correlations) that exist within the data. Unlike many areas of "hard" science with fixed well defined, formula-based relationships, most of the correlations that exist in Census data are poor. These correlations reflect a lot of the "trash" in the data -- rounding, indifferent responses/guessing, decimal errors, etc. A key EDA technique taught by the author is to use the typical/majority responses (and the understanding gained from an interactive analysis of all of the variables in the survey) to help us deal with the trash (or invented/indifferent responses -- "liars").

In some areas of the Bureau, Production Control Charts (PPCs) are now automatically produced as part of an enhancement to the user interface. Using the best X/Y correlations of the variables in our survey data sets, PPCs can be created -- including editing aids like the 80% confidence/outlier editing lines shown on the left below. Analysts can quickly now spot points like # 211 and # 150 as outliers. In most situations where these relationships are straightforward it is a simple matter to then edit these data.

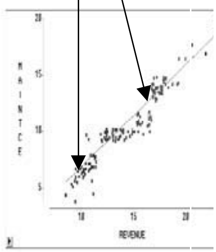
Even simple PPCs like this often reveal serious problems. In the example shown on the right below, there are serious problems with a subset of companies in the Bureau's Transportation Survey data. In this graph, the highlighted points (with unusually high Revenue and low Shipping Weight) are simply out of place). (Note: the units shown are the log values of these data.) Even a non-statistician can easily detect that there is a problem with this "fit". Further investigation disclosed that these companies were reporting Shipping Weight in their typical unit of measurement -- tons --instead of the unit of measure requested on the questionnaire -- pounds. This problem had been going on for years. Given their lower values, the conventional/fixed methods of editing had not detected the problem. The fixed methods were "blind" to the unforeseen anomalies in the data -- unusual points that are easily determined with the proper graphical views.

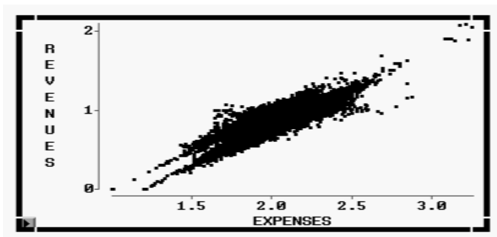




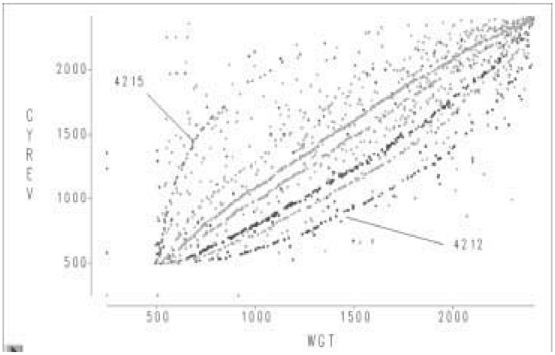
2.2. Examples 2 and 3 – Using Basic Editing Techniques to Find Previously Undetected Errors/Patterns

In some areas of the Bureau, even the simplest graphical EDA techniques have often proved to be revolutionary. Our EDA course stresses routine plots of the relationships between ALL of the survey -- not just the best correlations. (These plots are supplemented with a number of EDA techniques -- like brushing between the points in these graphs.) These simple graphical methods offer key insights into many potential problems in the data. For instance, humans can easily see a distinct clustering of the points in the graph on the left. Many computer algorithms would simply ignore these successive levels of point cloud clusters -- and report a good linear relationship between reported Revenues and Maintenance Costs (Expenses) for this industry group. In like manner, the graph on the right below illustrates an example of a problem that had gone undetected until a review using graphical methods. We can easily see that there are two rather distinct clusters evident in these data. Because of a flagging error, companies that were tax exempt were accidentally intermixed with companies that were not tax exempt. By graphing the data (and brushing on the flag variable), the analysts were able to easily detect a problem with the flags in the data that our "blind" editing software did not detect.





Thus, just plain exploration is a key aspect of these EDA graphical methods. Interactive EDA methods not only offer us a global perspective of the data -- but also the ability to explore/understand the relationship between many more (less correlated) variables than are typically examined with our traditional methodology. These perspectives can also lead to the discovery of problems actually generated by our traditional/fixed ways of editing/imputing data. At the Bureau, the graph below is now commonly referred to as a "football graph" (see DesJardins 1998). During one EDA session, while routinely plotting/scanning out data, this unusual pattern in data literally jumped out at us. The plot shows the reported Revenues vs. Shipping Weight (ranked by size) for all of the companies in our Transportation Survey. Hopefully, slight variations in the tone of the points in this graph are apparent. These reflect the distinct lines that were shown on our CPU screen in a different color. These colors/lines represent each of the different SIC codes assigned to our Transportation Survey -- for instance: "long distance shipping with storage, "short distance without storage", etc. On our CPU screen, in color, it was a quite spectacular graph! Further investigation showed that each of these lines was actually created by our imputation methodology. (If a Shipping Weight was not reported, then the reported Revenue was used to assign a Weight. In the first days of this preliminary survey, a large number of points needed to be imputed.) The other, rather randomly distributed, points in this graph are companies who reported both Revenue and Weight. Clearly, these imputation formulas -- which are fixed a priori -- caused us to over-impute Weight values for SIC # 4213 and under-impute Weight values for SIC code # 4212.

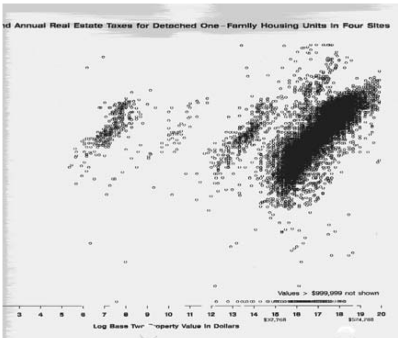


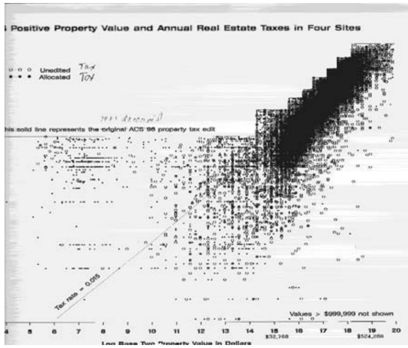
2.3. Example 4 - The Next Data Editing Step -- Using Pattern Graphs for Detecting Common Errors

The graph shown below is an example the "common errors" curriculum that is part of an EDA class taught by the author. The real goal of this EDA class is to teach an efficient (non labor-intensive) ***visual methodology*** that will help the analyst quickly detect errors and understand their data. Plotting examples of common errors trains the eyes of analysts to recognize typical error patterns. The analyst can then quickly recognize those patterns in their surveys -- quickly isolating points that are valid and those that simply do not make sense.

Again, fixed methods of editing can often be “blind” to some of the errors in data. This example shows how the application of these fixed methods (that had not been reviewed carefully) can massacre data. Below on the left we see a scatterplot of taxes paid (the Y variable) versus the corresponding assessed property values (the X variable) for a number of homes in a county in the USA (again, the data is in log values). As can be seen, the data all fall within a rather well defined range of Y values (taxes paid). However, we can see successive smaller clusters of "outlier" points for successively lower and lower levels of property values. Again, note that these successively lower property values all fall within this same range of taxes paid. Something is clearly wrong here. On the right we see the results of one of our old (blind) editing techniques -- as it tried to account for these smaller clusters of outlier data to the left of the main point cloud of these data. (Note: In the graph

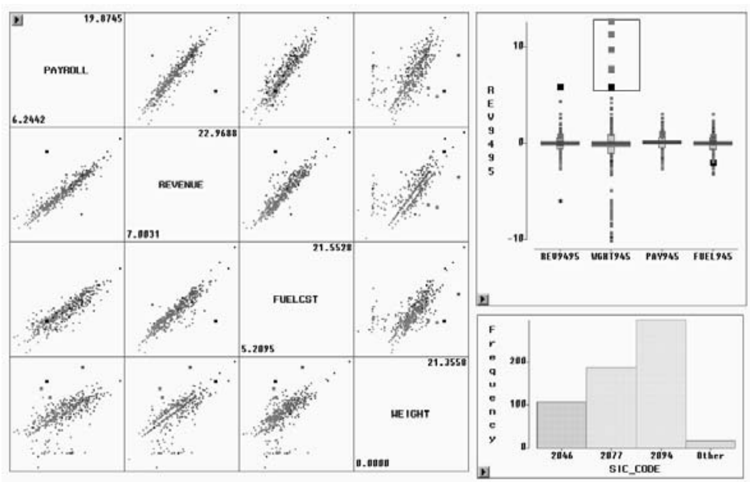
on the right, the fit line reflects the country tax rate -- and the darker/filled points are values imputed by this algorithm.)
A common error in recording or transcribing data is dropping zeros -- showing, for instance, a property valued at $130,000 as $13,000. If we look closely at the plot on the left, we can see that these outlier clusters actually represent errors of 10's, 100's, and 1000's. Note how the blind algorithm took "bites" out of successive steps of "good" values as it tried to adjust these data for these outlier points!





2.4 Example 5 - Production Work -- DP, IP Plots, and EDA Mapping

Perhaps the best example of the use of "live" graphs and interactive EDA methodology for day-to-day production editing would be represented by the Data Profile (DP) Plot. We currently use the DP plot (below) in conjunction with a number of interactive EDA techniques (like brushing, animation, juxtaposition, and the assignment of a progressive color scheme). Using this plot, analyst can quickly conduct a multivariate analysis of ***all*** of the key variables from a survey form -- and gain a real understanding of their data.



(Likewise, special "Industry Profile Plots" have also been devised by the author to maximize the productivity of analysts. Unfortunately, the paper length and limitations of this medium does not allow for explaining these techniques more fully here.)

On a much higher plane, contextual editing is a more advanced (but fundamental) EDA methodology that is stressed in the author's EDA course. Plotting our data against a geographic reference base gives our analysts a methodology to reflect the basic framework from which these data are gathered. To this end, new software has been devised that allows a marriage between our EDA software and ESRI Corporation's mapping package. With this new software capability -- which markedly exceeds the capabilities of most geographic information system (GIS) packages -- exciting things are happening!

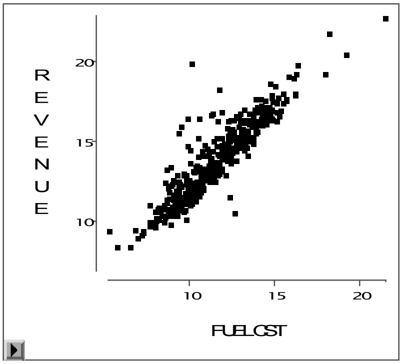
2.5. Example 6 - Advanced Data Editing Methodology -- Leverage Plots/Inliers

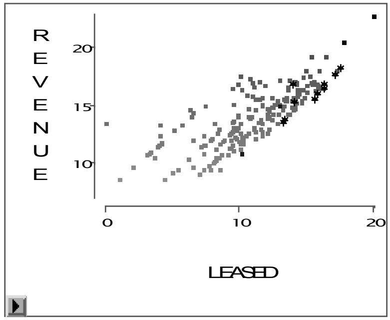
Leverage plots are another very helpful EDA tool. They have been particularly helpful in identifying a big problem with Census data -- hidden inliers. The partial leverage plot for each explanatory variable is used as an indicator of the relative influence of each observation on the parameter estimates. For a given explanatory variable, the partial leverage plot is the plot of the response variable and the explanatory variable after they have been made orthogonal to the other explanatory variables in the model.

For linear models, the partial leverage plot for a selected explanatory variable can be obtained by plotting the residuals for the response variable against the residuals for the selected explanatory variable. The residuals for the response variable are calculated from a model having the selected explanatory variable omitted, and the residuals for the selected explanatory variable are calculated from a model where the selected explanatory variable is regressed on the remaining explanatory variables.). Let ry[j] and rx[j] be the residuals that result from regressing y and X(j) on X[j]. Then a partial leverage plot is a scatter plot of ry[j] against rx[j].

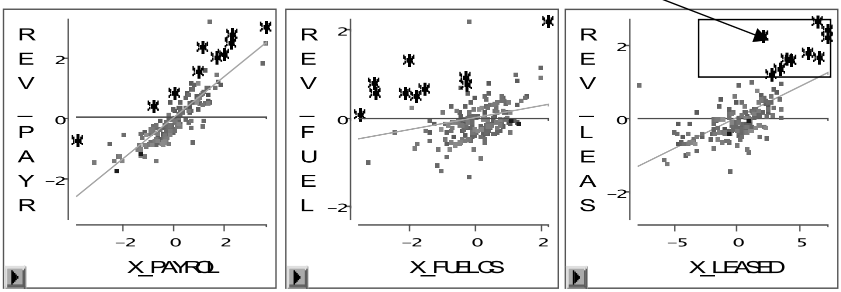
Two reference lines are displayed in the plot. One is the horizontal line of Y=0, and the other is the fitted regression of ry(j) against rx(j). The latter has an intercept of zero and a slope equal to the parameter estimate associated with the explanatory variable in the model. The leverage plot shows the changes in the residuals for the model with and without the explanatory variable. For a given data point in the plot, its residual without the explanatory variable is the vertical distance between the point and the horizontal line; its residual with the explanatory variable is the vertical distance between the point and the fitted line.

Below we see another two scatterplots of our trucking company data -- reported revenues and reported fuel costs -- and revenues and reported leasing costs. Aside from a few outlier responses and the fact that the lease cost data is a lot less correlated with revenues (than fuel costs), no alarm bells ring when we first look at these graphs.

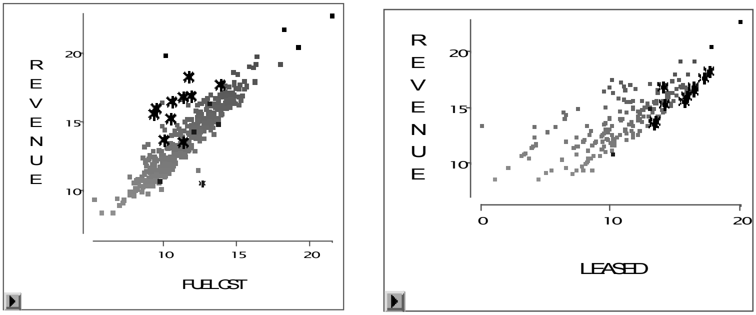




Below we see leverage plots of a fit of the reported revenue versus three variables from this survey. One needs little training to understand the basics of leverage plots. The graph on the left (for payroll) represents a good fit -- with a fit line close to 45 degrees and the points rather well distributed along the line. The graph in the center (for fuel costs) represents a poor fit -- with a rather random distribution of points and a leverage fit line close to the null line. If the brushed points within the square in the graph on the right (for lease costs) were eliminated from the fit calculations, this graph would more closely resemble the ("good") payroll graph on the left. Thus, we begin to suspect something about this subset of very suspicious ("outlier") points (companies) in the lease cost plot. Please note how interactive brushing also shows these companies have the highest values in the leverage plots of payroll and fuel costs -- very suspicious!



Returning now to the original scatterplots of these data (below), we can see that these points show up as a special cluster of points in the revenue versus fuel costs data (those marked with a dark "*"). These points stand out as a "bump" -- of companies with higher revenues for lower fuel costs than the majority of these other companies. How do these companies achieve higher revenues using less fuel -- what is their secret? Please note that, in the corresponding scatterplot of revenues and leasing costs, most of these points are clearly inliers. They only discretely show up along the lowest edge of these data. For each of these companies, the lowest levels of revenues correspond to the highest levels of lease costs. These companies have (higher cost) leasing contracts -- which would provide trucks that come with a driver and fuel. This helps us understand the how those companies attain higher revenues with lower fuel costs.



As can be seen, our EDA classes gives analysts (subject matter specialists) new tools that promote a better understanding of their data. These lease cost anomalies would certainly affect the imputation methodology that is used. These new tools not only quickly isolate data that may be in error, but offer an opportunity to truly understand it -- to show us data/companies (that needs to be treated differently) that are hidden to traditional statistical methodology.

2.6. Summary

The previous section covered powerful EDA methods that can be easily applied by subject matter specialists who are analyzing and correcting survey data. In addition to help in identifying outliers and inliers (and just plain liars), we have found EDA techniques to be a very helpful in the following areas: (1) data analysis such as modeling and identifying clusters/unique subsets, (2) checking basic methodology such as imputation and sample weighting, and (3) really understanding our data -- multivariate analysis.

In summary, the Census Bureau is entering a whole new world of data analysis capability. This is made possible by new, very fast hardware (i.e., Pentiums and Unix Workstations) and powerful, easy to use, point-and-click software (JMP and INSIGHT from SAS Institute). Formerly, analysts had to learn the intricacies of programming or often have to wait for months for systems development efforts to produce custom software that they needed for their data analysis tasks. Instead, in conjunction with a quick, 40 hour, EDA course taught by the author, analysts are being taught a variety of powerful interactive EDA techniques using this easy to learn (basically point-and-click) software. The design of this courseware is revolutionary in two other ways as well. First, it stresses a multivariate analysis -- of all of the variables on the survey form -- allowing for comparisons between variables in these data sets that, until now, had not been compared.

The result is that we gain a real understanding of these data. Second, it is designed for subject matter specialists who have only a moderate statistical background -- to give them key insights into their data.

3. Concluding remarks

The ability of the eye-mind combination to discern subtle and complicated relationships during review of graphs has long been known. This paper has demonstrated that is it possible to use new, powerful, user-friendly graphical software to explore and correct data. Through his 40 hour EDA/Graphics course, the author has been particularly effective in moving the ideas in to day-to-day practice. The use of these EDA/Graphics techniques is also now a part of mainstream statistical methods for exploring and reviewing complicated statistical models.

References

Cleveland, William (1993), *Visualizing Data*, Hobart Press: Summit, NJ

DesJardins, David (1998), New Graphical Techniques for the Analysis of Census Data, *Statistics Canada Conference Proceedings.*

Granquist, L. (1997), Macro-Editing - The Aggregate Method, *Statistical Data Editing, UN Conference of European Statisticians Statistical Standards and Studies, Geneva (Switzerland).*

Hogan, Howard (1995), How Exploratory Data Analysis is Improving the Way We Collect Business Statistics, *American Statistical Association, Proceedings of the Section on Survey Research Methods,* 102-107.

Winkler, W. E. (1997), Problems with Inliers, *paper presented at the European Conference of Statisticians, October 14-17, 1997, Prague, Czech Republic*.

A Generalized Edit and Analysis System for Agricultural Data

Dale Atkinson, Carol House
National Agricultural Statistics Service, U.S. Department of Agriculture
3251 Old Lee Highway, Fairfax, VA 22030-1504, USA
e-mail: datkinson@nass.usda.gov, chouse@nass.usda.gov

Keywords: Edit, Analysis, Integrated, Interactive, Seamless, Modular design.

1. Background

In 1997 the responsibility for the quinquennial census of agriculture was transferred from the U.S. Bureau of the Census (BOC) to the National Agricultural Statistics Service (NASS) in the U.S. Department of Agriculture. This fulfilled a goal of NASS to become the national source of all essential statistics related to U.S. agriculture. It also provided an opportunity for the Agency to improve both the census and its ongoing survey and estimation program through effective integration of the two.

The timing of the transfer, however, severely limited the changes NASS could make for the 1997 Census of Agriculture. To complete this census NASS formed a Census Division that had primary responsibility for managing the day-to-day operations of the census activities. This Division included former BOC employees who transferred to NASS with the census. Much of the data collection, data capture and editing was contracted out to the BOC's National Processing Center (NPC) in Jeffersonville, Indiana, which had also assumed these functions in prior censuses.

NASS *was* able to make significant changes in some of the census processes. Specifically, the Agency was able to utilize its 45 State Statistical Offices (SSOs) in coordinating census data collection with that of its ongoing survey program. The SSOs also played a key role in the processes from macro-level editing through the final review of the data for publication. In previous censuses these processes had been centralized and the States' data were reviewed sequentially, in a pre-determined order. By decentralizing the review process, the States' data were reviewed concurrently -- significantly reducing the time from initial data aggregation to publication. This allowed the publication of 1997 census data a year earlier than those of previous censuses.

However, some of the main benefits of NASS acquiring the census of agriculture have yet to be realized. In particular, a proper integration of the census program with NASS' traditional program figures to improve the quality and efficiency of each. These are benefits that Agency management has targeted for 2002 and beyond. To begin the process of integrating the programs NASS took two major steps. The first of these was the creation in late 1998 of the Project to Reengineer and Integrate Statistical Methods (PRISM). The team appointed to manage this project was charged with conducting a comprehensive review of all aspects of the NASS statistical program and recommending any changes needed. The second step was a major structural reorganization of the Agency. This reorganization essentially absorbed the staff and functions of the Census Division, as formed for the 1997 census, into an enhanced survey/census functional structure. The reorganization was designed to increase efficiency and eliminate duplication of effort by integrating census responsibilities throughout the structure.

2. Introduction

The census processing system needed to be reengineered prior to 2002. With the transfer of census responsibility in 1997, NASS had inherited an aging system that had been used, largely unmodified, since 1982. It was out-of-date technology-wise and, to a lesser extent, methodology-wise. The system was relatively inflexible in that decision logic tables (DLTs) were "hard coded" in Fortran. It was programmed to run on aging DEC VAX machines running the VMS operating system. While manual review and correction could be performed on standard PC screens, some functionality was lost when the system was used with display terminals other than the amber-screened DEC terminals for which it was designed. In general, the record review and correction process at both the micro- and macro-levels involved navigating an often-frustrating combination of function and control keys. The system had served its purpose through the processing of the 1997 census, but it was time for a more up-to-date system.
In September 1999 the Processing Methodology Sub-Team of PRISM was chartered to specify a new edit, imputation and analysis system for the 2002 Census of Agriculture and subsequent, large NASS surveys. This group reviewed editing literature and processing systems used in NASS and other organizations (U.S. Bureau of the Census, 1996 and Weir, 1996) to synthesize the best of what was available into its recommendations for the new system. In February it published its findings and recommendations in an internal Agency research report. The report highlighted the team's guiding principles, as follows:

1) Automate as much as possible, minimizing required manual intervention – Having dealt exclusively with much smaller sample surveys in the past, the NASS culture has been to touch every questionnaire and have statisticians manually specify needed data changes in response to automated edit flags. The sheer volume of data precludes this option for the census and necessitates a system that makes more editing/imputation decisions automatically, without manual intervention

2) Adopt a "less is more" philosophy to editing – There's a tendency in many organizations to over-edit data -- automatically and/or manually. A leaner edit that focuses on critical data problems is less resource intensive and often more effective than a more complex one.

3) Identify real data and edit problems as early as possible -- One of the concerns about the edit used for the 1997 census was that SSO analysts had nothing to review from the highly automated process for several months after editing started. Except for a few who were temporarily detailed to NPC to correct edit failures, SSO statisticians were unable to see the data until they were weighted for nonresponse and aggregated. This was often six months after initial data collection. The delay caused problems that could have been more effectively handled earlier in the process and imposed additional stress on the SSOs by complicating and compressing their data review time

4) Design a system that works seamlessly – While 'seamless' means different things to different people, what is needed is a system in which all the components interrelate smoothly such that the analyst can quickly and easily navigate to any screen and get any auxiliary data needed to identify and resolve a data problem. A system is definitely not seamless if the user has to log into various computer systems separately to obtain needed auxiliary data or run an ad hoc query. Lack of 'seamlessness' was a problem that reduced the effectiveness of the 1997 census processing system.

5) Use the best features of existing products in developing the new system -- By the time the 1997 Census of Agriculture was completely put to rest, the 2002 Census of Agriculture was uncomfortably close at hand. The short developmental time would preclude "re-inventing the wheel." It was imperative that NASS incorporate the best aspects of what had already been done research-wise and developmentally in NASS and other organizations to expedite the process as much as possible.

In view of the above guiding principles the sub-team documented the features it felt the new system should include (Processing Methodology Sub-Team, 2000). Considerable emphasis was placed on minimizing unnecessary review and on the visual display of data. The sub-team discussed display attributes and methodologies that could be used to identify problematic data with high potential impact on published estimates. The 'features' section of their paper discussed the issue of refreshing the review screens as error corrections are made and stressed the need for the system to help manage the review process (i.e., to identify records that had already been reviewed, through color and/or special characters). The sub-team concluded its paper with the following recommendations:

i) To the extent possible, use Fellegi-Holt methodology in the new system.
ii) Have the computer automatically correct everything with imputation at the micro-level (i.e., eliminate the requirement for manual review).
iii) Utilize the NASS data warehouse as the primary repository of historical data and ensure that it is directly accessible by all modules of the new system.
iv) Design the system with tracking and diagnostic capabilities to enable the monitoring of the effect of editing and imputation. Develop analytics for a quality assurance program to ensure edited/imputed data are trusted.
v) Incorporate a score function to prioritize manual review.
vi) Provide universal access to data and program execution within the Agency.
vii) Ensure that the system is integrated into the Agency's overall information technology architecture.
viii) Make the system generalized enough, through modular design, to work over the entire scope of the Agency's survey and census programs.
ix) Enable users to enter and access comments anywhere in the system.
x) Present as much pertinent information as possible on each screen of the system and provide on-screen help for system navigation.
xi) Consider the use of browser and Java programming technology to assist in integrating parts of the system across software, hardware, and functions.
xii) Designate a developmental team to take this report, develop detailed specifications and begin programming the system.

3. The System Development

In response to recommendation *xii*, a number of working groups were formed to focus on various aspects of the processing system development. These included groups addressing check-in, data capture, edit specifications, interactive data review (IDR) screens, imputation, analysis, and census coverage evaluation. In order to ensure consistency of decisions across the working groups in assembling the system an oversight and technical decision-making body, the Processing Sub-Team, was formed of the leaders of the individual working groups. This sub-team was charged with considering the overall system flow and ensuring that the individual modules work together effectively. The sub-team members keep each other informed about the activities of their individual groups, thus ensuring consistency and that required connectivity is addressed. The sub-team also serves as the technical decision-making body for crosscutting decisions that can't be made by the individual working groups. The following sections describe plans for selected modules of the system, the progress made to date and some key issues that the working groups are grappling with.

3.1 Data Capture

As was the case in 1997, NASS will contract the printing, mailing and check-in of questionnaires and the data capture activities to NPC. While all data capture for the 1997 Census of Agriculture was accomplished through key-entry, NASS' early discussions in preparing for 2002 indicated that scanning could be used to capture both an image of the questionnaire for interactive data review and the data itself, through optical/intelligent character recognition (OCR/ICR). Preliminary testing done with the Agency's Retail Seed Survey supported the practicality of using scanning for data capture. Testing of the OCR/ICR process for this survey was conducted at three different confidence levels (65, 75 and 85%). The outcome of this small test was that at 65%, 4% of the characters were questionable; at 75%, 5-7%; and at 85%, 13%.

NASS will utilize scanning with OCR/ICR as the primary mode of data capture for 2002. Current plans are to start with the industry standard confidence level of 85%, but this might be adjusted with further experience in using the system with agricultural census data. Results from the recently completed census of agriculture content test should help fine-tune the process. Questionable returns will be reviewed, with erroneous data re-entered by correct-from-image (CFI) key-entry operators. The scanning process will produce data and image files, which will be sent to the Agency's leased mainframe computers at the National Information Technology Center (NITC) in Kansas City, Missouri for further processing. The data will pass into the editing system and the images will be brought into the interactive data review screens that will be activated from the analysis system to review and correct problematic data.

3.2 Edit

As the edit groups began to meet on a regular basis the magnitude of the task of developing a new editing system became obvious. The machine edit/imputation used for the 1997 census was enormous. It had consisted of 54 sequentially run modules of approximately 50,000 lines of Fortran code, and the sheer volume of the input decision logic tables (DLTs) was staggering. Through 1997, the census questionnaires had changed very little from one census to the next, so the DLTs and Fortran code had required little modification. For 2002, however, an entirely new processing system would be built on a questionnaire that was also undergoing radical changes. Some of the questionnaire changes were necessitated by recent structural changes in agricultural production and marketing, while others were due to the planned use of OCR/ICR for data capture. In any case, the group members were saddled with the onerous task of working through the mountainous DLTs from 1997 to determine what routines were still applicable and, of these, which should be included in the new system specifications. One of the key edit issues is reducing manual review without damaging data quality. In processing the 1997 census data, the complex edit corrected all critical errors and the staff at Jeffersonville manually reviewed ALL "warning" errors. The approximate workload and time invested in this activity follows:

- Approximately 1,800,000 records passed through the edit at least once. Of these, 470,000 (26%) were flagged with warning errors. About 200,000 (47%) of the flagged records required updates.
- About 4,000 staff days were spent performing the review in Jeffersonville

For 2002, the edit review (and analysis) will be performed in NASS' SSOs. Considering the expected staff shortages in 2002 relative to 1997, the above figures would represent an intolerable commitment of staff resources. Furthermore, indications are that this amount of manual review is not altogether needed or (in some cases) desirable. Table 1 shows the relative impact of 1) the automatic (computer) edit changes with no manual review; 2) edit changes with/from manual review and 3) changes made during analytic review. Due to deficiencies in the edit coding, some changes made totally by computer could not be cleanly broken out from those with manual intervention, resulting in an overstatement of the manual edit effect. All changes made during analytic review resulted from human interaction and are considered part of the impact of manual review.

Table 1 shows that the overall effect of the edit/imputation/analysis process was relatively small for most items, especially crop acreages. Considerably larger adjustments are required for both nonresponse and undercoverage. While admittedly these numbers only reflect the impact on high-level aggregates (U.S. level) and the processing can often be more beneficial at lower levels (e.g., county totals), the size of the adjustments still raises questions about the efficacy of the extremely resource-intensive data editing and review process. Such considerations underpinned our two guiding principles of 1) adopting a "less is more" philosophy to editing and 2) automating as much as possible.

Table1: *Relative Impact of the Editing/Imputation/Analysis Processing of the 1997 Census of Agriculture Data (U.S.* Level*)*

Characteristic	Net Effect of Automated Edit Changes (%)	Net Effect of Edit Manual Review (%)	Net Effect of Analytic Review (%)	Total Effect (%)	Total Manual Effect (%)
Corn Acres	(0.24)	(3.97)	0.26	(3.94)	(3.71)
Soybean Acres	(0.20)	(2.33)	0.31	(2.22)	(2.02)
Wheat Acres	(0.69)	(4.18)	(0.01)	(4.88)	(4.19)
Cotton Acres	(0.10)	(0.29)	(0.27)	(0.66)	(0.56)
Cranberry Acres	0.13	1.72	(4.04)	(2.18)	(2.32)
No. of Cattle	0.74	4.75	(0.74)	4.74	4.01
No. of Hogs	0.17	(4.23)	(3.92)	(7.98)	(8.15)

3.3 Imputation

Certainly one of the key considerations in moving to an automated system is determining how to impute for missing and erroneous data. The imputation group is currently working through the census questionnaire question-by-question and section-by-section to determine the most effective routines to use. Nearest-neighbor donor imputation will play a strong role in filling data gaps. The group is currently developing a SAS®-based donor imputation module, which will provide near optimal imputations in certain situations where high quality matching variables are available. The group will be leveraging the Agency's relatively new data warehouse capabilities of providing previously reported survey data. The data warehouse was populated with the 1997 census data and contains the data from most of the Agency's surveys since 1997. As such, it serves as a valuable input into the imputation process, since many of the respondents in the current survey will have responded to one or more prior surveys. The warehouse data can provide direct imputations in some cases and identify items requiring imputation in many others.

A review of the imputation done for the 1997 Census of Agriculture and the current plans for 2002 indicates the vast majority of the imputation performed will be deterministic (e.g., forcing subparts to equal a total). Deterministic imputation could amount to 70-80% of all imputation for the 2002 Census of Agriculture. Nearest neighbor donor imputation will likely account for 10-20%, while direct imputation of historical data, perhaps 5-10%.

3.4 Analysis

3.4.1 General Description

The analysis system is perhaps the module of interest to the broadest audience in NASS. This module will provide the tools and functionality through which analysts in Headquarters and our SSOs will interact with the data. All processes prior to this point are ones with no manual intervention or, in the case of data capture, one in which only a few will touch the data. As one of our senior executives aptly put it, "All this other stuff – data capture, edit and imputation will happen while I'm sleeping. I'm interested in what will go on when my eyes are open." That's analysis!

Because of the broad interest in and the expected large number of users of the analysis system, the development team has made a special effort to solicit user input into its specification. The working group chartered to design and program this module circulated a hard-copy prototype of the proposed system to staff throughout the Agency early this year. This exercise resulted in very useful feedback from potential users. The feedback received has been subsequently worked into the module specifications.

3.4.2 Micro-Analysis

After the data have been processed through the edit and imputation steps, during which essentially all critical errors have been computer corrected, they are ready for SSO review in the Analysis System. The first of two analysis phases, micro-analysis, begins immediately. During micro-analysis SSOs will review (and update, if necessary) all records for which imputation was unsuccessful, all records failing consistency checks, and all those with specific items that were flagged for mandatory review. Such records are said to contain critical errors and must be corrected. This work will be done while data collection is ongoing, and will allow ample time for any follow-up deemed necessary. As review time permits, the system will also provide the capability to review records that have no critical errors, but may be nonetheless of concern. These would include those identified by the computer as influential or high scoring or with potential problems identified through graphical analytic views. Unlike the 1997 edit, warning errors will NOT be automatically corrected nor require manual intervention.

A score function is being developed for 2002 Census of Agriculture to ensure that the records manually reviewed are those that are expected to have a substantial impact on aggregate totals. The quality of county aggregates is of particular concern with the census of agriculture. Therefore, the score function used for 2002 will be one that assigns high scores to records whose current report for selected characteristics represents a large percentage of the previous census' county total for that characteristic.

Micro-level graphics are simply a collection of record level information shown together for all records for specific item(s) of interest. The user will have the option of sub-setting the graph by selecting a group of points or by specifying a sub-setting condition. For some plots, the option of additional grouping and/or sub-grouping of a variable(s) through the use of colors and symbols will be available (e.g., by size of farm, type of operation, race, total value of production, other size groups). Scatter plots, box-plots and frequency bar charts of various types will be provided. All graphics will provide drill-down capability to data values and the IDR screens to review and update problematic records.

Finally, the system will track IDs that have been previously reviewed, compare current values to historic data, allow for canned and ad hoc queries and have a comments feature to document actions. Micro-analysis will also include tables to review previously reported data for non-responding units. This will allow SSOs to focus nonresponse follow-up efforts on the most "important" records.

3.4.3 Macro-Analysis

The second phase of analysis, macro analysis, begins immediately after preliminary weighting (adjusting for under-coverage and non-response). Macro-analysis uses tables and graphs to review data totals and farm counts by item, county and state. While the macro-analysis tools will retain the key objectives of the analytical review system used for the 1997 census, it will be much more interactive and user-friendly. The focal point of the macro-analysis will be a collection of graphics showing aggregate data at state and county levels. These graphics will include dot plots or bar charts of county rankings with historic comparisons, state maps with counties color-coded by various statistics and scatter plots of current vs. previous data.

The new macro-analysis tool will also be integrated more effectively with the Agency's data warehouse and its associated standard tools for user-defined ad-hoc queries. Graphics or tables will be used to compare current census weighted totals and farm counts against previous census values and other published estimates. There will be a prepared library of database queries, in addition to the ability to build your own. Analysts will drill down to the IDR screens to verify/update records. If micro-analysis is done effectively, the number of issues to be dealt with in this phase will be fewer than in 1997, when no micro-analysis module was available.

The macro-edit can be run as soon as data collection is complete, the last records are run through edit and imputation, and preliminary weights are available. The objective of the macro review will be the same as for 1997. That is, an analyst will be responsible for the complete review of all the state and county totals. According to a state's particular needs and characteristics the SSO's managers can elect to either 1) assign an analyst to a county for the review of all items, 2) have a commodity specialist review items by state and county, or 3) use a combination of both. In any case, every item in each county must be reviewed, and a check-off system will be provided in the analysis system to ensure this is achieved.

4. Development Status

Timelines have been developed for specification and development of the various modules and the groups are working hard to stick to them. Due to a number of factors beyond their control the developmental work started at least a year later than it should have, considering the magnitude of the system overhaul. In spite of the delays and overall staff shortages as compared to what was available for past censuses, the groups have done a fantastic job of moving ahead with the developmental work.

5. Issues

One of the key issues in edit development is determining what edits are essential to ensure the integrity of the data without over-editing. This is something that the edit group and Processing Sub-team have struggled with. The team members represent an interesting blend of cultures. The longer-term, pre-census NASS staff developed within a culture of processing the returns from its sample surveys, where every questionnaire is hand-reviewed and corrected as necessary. While there is a need for some of this extra attention for sample surveys since survey weights can be high, this type of approach is nonetheless prone to manual over-editing. The NASS staff that came over with the census are much more comfortable with automatic editing/imputation and are perhaps overly interested in having the system account for every possible data anomaly. This approach can lead to an excessively complex system that automatically over-edits data.

The combination of these two cultures has resulted in some interesting discussions and decisions relative to the guiding principles of automating as much as possible and adopting a 'less is more' philosophy of editing. Everyone has his or her own pet anecdote indicating a "crucial" situation that a reduced edit would not identify and correct. Such concerns have resulted in some compromises in the editing approach taken for 2002. The new processing system currently being developed will look more like a SAS version of the 1997 edit than the greatly reduced, predominantly error-localization driven system envisioned by the Processing Methodology Sub-Team. The bottom line for 2002 is that there will be 49 DLT edit modules, which will consist of much of the same type of intensive, sequential "if-then" edit conditions that existed in 1997. There are some notable differences in the processes, however. There will be a presence of GEIS-type error-localization (Statistics Canada, 1998) in the new system in addition to the 1997 style editing. Imputation has been moved out of the edit DLTs to a separate module of the system. This module will make strong use of nearest-neighbor donor imputation, enhanced by previously reported data from the Agency's data warehouse. The error-localization presence will help ensure that the imputations will pass all edits. The approach to be used in 2002 will serve as a foothold for the approach (Fellegi-Holt, 1976) initially endorsed by the Processing Methodology Sub-Team. For 2007 there will be a strong push to simplify the edit and increase the error-localization presence or move to the NIM-type approach of Statistics Canada (Bankier, 1999).

Another key issue in assembling the system lies in how much modularity/generality is possible. All efforts currently are, and need to be, directed at having a system in place and tested for the 2002 Census of Agriculture. Due to the tight time frame the developmental team is working with, some compromise on the goal of generality is inevitable. The evolving system is being developed modularly, however; so some retrofitting of generality should be possible.

One of the questions that have yet to be answered is whether runtimes and response times will be adequate? The current plans for the processing system are complex, requiring considerable cycling through the various sections of the questionnaire. Whether or not the runtimes on batch aspects of the system and response times on the interactive portions will be within workable tolerances will not be fully known until more of the system is built. If the answer in either case turns out to be negative, short-cuts will need to be taken to make the system workable.

Exactly what combination of processing platforms will be used in the final system is another issue that has yet to be fully decided. It will be comprised of some combination of the Agency's leased mainframe, its UNIX boxes and its Windows 98 machines on a Novell wide-area network. Since the system is being written in SAS, which will run on any of the three platforms, the processing platform decision has been delayed up to now. However, in the interest of seamless interoperability it will need to be made soon.

References

Bankier, M. (1999), Experience with the New Imputation Methodology Used in the 1996 Canadian Census with Extensions for Future Censuses, *Conference of European Statisticians, June 2-4, 1999, Rome, Italy*.

Fellegi, I.P. and Holt, D. (1976), A Systematic Approach to Automatic Edit and Imputation, *Journal of the American Statistical Association*, March 1976, Vol. 71, No. 353, pp. 17-35.

Granquist, L. and Kovar, J. (1997), Editing of Survey Data: How Much is Enough? in L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), *Survey Measurement and Process Quality*. New York: John Wiley and Sons.

Luzi, O. and Pallara, A. (1999), Combining Macroediting and Selective Editing to Detect Influential Observations in a Cross-Sectional Survey Data, *Conference of European Statisticians, June 2-4, 1999, Rome, Italy*.

Processing Methodology Sub-Team (2000), Developing a State of the Art Editing, Imputation and Analysis System for the 2002 Agricultural Census and Beyond, NASS Staff Report, February 2000.

Statistics Canada (1998), Functional Description of the Generalized Edit and Imputation System, Technical report from Statistics Canada.

Todaro, T.A. (1999), Overview and Evaluation of the AGGIES Automated Edit and Imputation System, *Conference of European Statisticians, June 2-4, 1999, Rome, Italy*.

U.S. Census Bureau (1996), StEPS: Concepts and Overview, Technical report from the U.S. Census Bureau.

Weir, P. (1996), Graphical Editing Analysis Query System (GEAQS), *Data Editing Workshop and Exposition, Statistical Policy Working Paper 25*, pp. 126-136, Statistical Policy Office, Office of Management and Budget.

7 June 2001

Invited Paper Session

AGRICULTURAL MODELLING

Chair: J. Karlsson

Data Needs for Modelling Sustainable Development in the Canadian Agricultural Sector

Robert J. MacGregor, David Culver, Robert Koroluk
Strategic Policy Branch, Agriculture and Agri-Canada
960 Carling Avenue, Building 74, Ottawa, Ontario, Canada, K1A 0C6
e-mail: macgrbo@em.agr.ca , culverd@em.agr.ca and korolur@em.agr.ca[1]

Abstract: The theme for the plenary sessions is "The Agricultural Situation in the New Millennium" and how it relates to developed and developing countries. Within this context the new policy challenges and the implications of these for data collection and analysis are discussed. One point is clear—the policy questions that governments are required to address are changing and increasing in complexity. To an ever greater extent, the analysis demanded must be comprehensive in nature, quantifying the direct and indirect linkages and feedbacks that exist, and responding to questions relative to the economic, social and environmental implications of governments taking action or not. Globalization and the rise of multinational policy mean that analysis can no longer be confined to one region, commodity, sector or country in most instances. As the policy development process demands more rigorous and comprehensive analysis, the demand for data that underpins this work also increases. Analysts of all disciplines are required to rely to an ever increasing extent on models due to the inherent complexity of what we are dealing with. These models are intensely data hungry, limited only by our increasing capacity to build and run larger and more complex models. The message is that there will be an almost insatiable demand for more data that will need to be internally coherent to facilitate the multi-disciplinary effort required. The question is how the existing system, with its many flaws and shortcomings, will satisfy this demand in the years to come.

Keywords: modelling, sustainable development, agriculture, environment, knowledge system.

1. Introduction

After a period of retrenchment during the late 1980s and most of the 1990s where the key concerns of most OECD governments related to eliminating structural deficits and going through a process of policy and program reform to sort out the types of activities governments needed to be involved in, OECD countries are now gearing up to address the emerging challenges of the 21st century. After a period of reform that often had significant structural and political costs associated with it, there is no desire to fall back to using the

---

[1] This paper contains the views of the authors and not necessarily those of Agriculture and Agri-Food Canada or the Government of Canada.

same policy instruments of the past. Governments want to rely more heavily on economic instruments that allow market forces to play a significantly bigger role in finding efficient responses to issues raised nationally and internationally. They also realize that globalization exists and that the ability to act independently is diminishing constantly. Governments are also taking a more horizontal approach to policy development, realizing that most issues are of a cross-cutting nature. It is heavy jargon to say the inter-connectivity of our economies cannot be ignored. The need for policy coherency, nationally and globally, is another way of expressing this new reality.

One factor that governments now realize is that during the past period of retrenchment, their public policy capacity was not only greatly diminished, but also subject to what some refer to as “rust-out” as new investments in data, knowledge, systems and human resources were not made. The aging public service is a concern in Canada, as it is in many other countries. At exactly the same time that the demands for policy relevant information are increasing, governments are faced with the need to reinvest, rejuvenate their workforce, and update the data, research and modelling systems that are the foundations of good analytical work and which form the cornerstone of public policy.

Within this broader context, this paper will focus on the data needs that sector modellers involved in policy analysis for sustainable development are identifying. Section 2 provides a quick overview of the types of models being used in the policy development process. Section 3 focuses on the limitations linked to data. Section 4 presents a discussion of the importance of this to the policy development process. Section 5 provides some comments related to future directions. It is beyond the scope of this paper to cover comprehensively the situation in all countries. The paper will draw heavily on our experience in Canada. The objective of the paper is to outline the challenges as seen by policy analysts who use data every day, and not necessarily to provide solutions.

2. Quantitative Models Employed to Assess Sustainable Development

Usually, one would start the discussion with trying to define “sustainable development.” However, we will not attempt to do so here except to state that, from an analytical point of view, economic, social and environmental considerations must be included. From a governance point of view, in most developed countries clear policy signals have been sent that all three elements must be considered in the policy development process. For example, in Canada the Commissioner for Sustainable Development sits in the Auditor-General’s Office, reporting directly to Parliament and the public the degree to which the federal government abides by the policy directions set out in its sustainable development commitments. The Commissioner is very much concerned with the policy development process and with ensuring that all aspects of sustainable development are appropriately incorporated into the decision-making process. Some of the analysis to support this process would be qualitative in nature, but more and more integrated, quantitative analysis is being demanded.

Bio-physical models: The first type of analytical models are bio-physical in nature and model various aspects of the physical and chemical interactions that proceed when production processes employ natural resources. In agriculture, soils, water, air and bio-diversity are all impacted during the production process by natural phenomena (e.g. climate) and man. This type starts with models that look at only one aspect of the production process (i.e. plant growth or wind erosion models), often at plot or field levels,[2] but also considers models that look at these processes at different hierarchical scales (Dumanski et al., 1998). Historically, these models have not entered the policy development process directly as they do not generate information readily usable by policy makers. Normally the information was only illustrative, showing what could be occurring in a field and not at a scale relevant to the policy process. This process of moving to higher hierarchical levels is limited when data on farming systems employed on specific landscapes is not available. This is unlike economic analysis, where the last ton of wheat is produced (using a specific sort of technology under particular climatic conditions matters) when trying to determine the bio-physical implications of that activity.
One real problem arises. When we do try to apply these models, often they do not perform well once taken out of the research station setting where they were developed. The mathematical relations built into them are based on research data often pertinent to a relatively small ecosystem. Once you move out of that ecosystem, the performance drops off. Experience with EPIC or CENTURY would confirm this point where the process of calibrating to different ecosystems is required. Calibration is a data intensive exercise in both the basic research data and in the data on the physical, climatic and technological characteristics of each new zone. Sometimes the necessary data are collected, but are not located within an information system that makes the data readily available.

Economic/behaviour models: The second type of model include those normally developed by economists. Being social scientists, economists go to great lengths to ensure that behavioural responses to key stimuli are captured within these models (the behaviour of consumers and firms, as well as the political economy where possible). Economists have a whole range of models available to them that deal with the entire hierarchical scale—from individuals to farm level, commodity, sector or regional models, right up to national sector models, national general equilibrium models and world models. On the supply side, the behavioural relationships built into the models depend on the technology used, often represented in a highly stylized manner. They depend on the costs and market prices that the producer or sector faces. On the demand side, a regional or national demand relationship is estimated or employed. This level of aggregation at the consumer level tends to work because it does not matter which consumers changed their consumption pattern to buy the last unit of product consumed as the impact on the market is assumed to be the same.

---

[2] This discussion will not deal with the “lab” models that deal with the processes within an individual plant or animal, but rather with those models that deal with how agricultural production systems interact with their environment in the field.

At Agriculture and Agri-Food Canada (AAFC) we employ farm level models, national sector models (CRAM, FARM), world models (PEM, AGLINK) and general equilibrium models (I/O, GTAP) to undertake policy analysis for the agricultural sector.[3] The potential to use micro-simulation models has recently been evaluated. Generally speaking, the data needs for models higher up the hierarchical scale can be met from traditional sources that correspond well to the traditional types of analysis undertaken by economists. However, at lower scales, data shortcomings become more serious. For example in CRAM, production is represented at the sub-provincial level but almost no cost of production data exists that relates the technology that farmers are using to the costs incurred in employing that technology in Canada at this time.[4] Years of government cutbacks have largely eliminated the collection of this type of information in most provinces. Lack of statistically representative samples should not be underestimated, and the availability in the future of traditional market data should not be taken for granted. With greater prevalence of vertically integrated market structures, publicly available information tends to decrease. As bulk commodity markets are replaced by segregated product markets, the private value of this type of information increases, as will the tendency to treat firm level data as a private versus a public good.

Integrated economic/environmental models: The third type of model links directly the bio-physical models and the behavioural economic models, starting at the farm level, but also at the regional level with attempts made at the national and the world levels (Lee and Lovejoy, 1991). Many times only one environmental factor may be considered. For policy purposes, a much broader perspective must be taken because policy makers must understand the impacts, feedbacks and adjustments over time, as well as the inherent trade-offs normally involved. All the data limitations noted above are also pertinent to this type of model. For example, four shortcomings that have been identified in Canada to undertake this type of analysis include:

- Lack of knowledge on what farming practices (technology packages) are associated with the landscapes on which they are practiced.
- Lack of knowledge of the cost structure associated with those technologies and the variability inherent from year to year.
- Lack of knowledge about the aggregation issues as analysis is carried out at higher scales in an effort to provide information useful for policy development purposes and therefore a lack of understanding of the statistical properties of the estimates derived (i.e. the confidence interval) (Antle and Just, 1992).

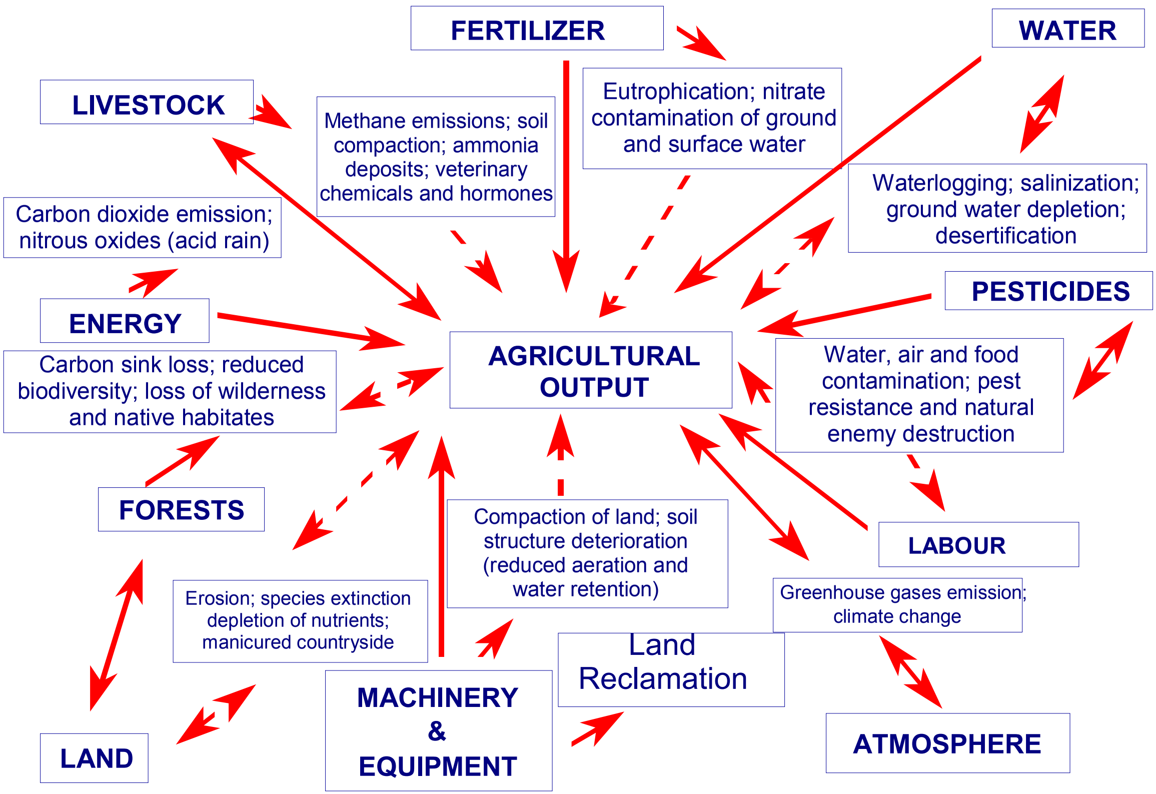
[3] CRAM—Canadian Regional Agricultural Model, FARM—Farm and Agriculture Regional Model, PEM—Policy Evaluation Model, I/O—Input/Output, and GTAP—Global Trade Analysis Project.

[4] The authors are currently working with Statistics Canada to determine if it is possible to use the Maximum Entropy methodology with farm records in the Whole Farm Data Base (Statistics Canada) to estimate cost functions at various scales.

- Lack of knowledge on the economic and bio-physical factors affecting the adoption of different technological packages.
-

Figure 1 provides an illustration of the complexity that exists in meeting the challenge that producers face, and in turn that we face as analysts. As analysts we must deal with this complexity for all types of producers with different technologies and social settings, commodities, ecosystems and at all hierarchical levels. Planned output is only one of a multitude of outcomes that result when resources and purchased inputs are employed by modern agriculture.

Figure 1: *Producer's Decision Making Framework*



Decision support models: The fourth type of model is called "decision support models" for lack of a better term. In this type of model, the weights that society puts on various outcomes need to be incorporated into the analysis. Public opinion surveys often supply the raw data for this type because invariably a valuation of non-market goods must be determined. Impacts on commodity markets are the stock and trade of agricultural economists. The value that society places on environmental and social outcomes is often classed as a public good or externality that is difficult to factor into the analysis.
Traditionally, dealing with the trade-offs and the choices that society collectively must take has been left to those directly involved in the political process as part of their mandate. However, this process is not very transparent as outcomes often depend on the ability of particular stakeholders to influence the decisions to obtain a favourable outcome for themselves, even if the decision mean a net welfare loss for society.

3. Data and Data Management System Needs for the 21st Century

Providing a linkage between environmental and other variables presents a particular challenge for data collection and for modelling sustainable development. Though the following discussion focuses on some of these challenges, the recommendations may be equally applicable to other areas suffering from data gaps.

Data sources for environmental modelling: The quality of the results of environmental models will depend to a large degree on the quality of the data from which the models are built. In most if not all countries, significant data gaps exist in the data required to build environmental models. In the absence of perfect data, researchers must consider all the potential sources of data and make the best use of each type. Four sources of data in each country could be considered: surveys, administrative data, remote sensing and research data
Surveys: If data are not available, researchers may have to consider collecting the data through a survey. In Canada, when national information on farm input management practices was not available, a specialized telephone survey was undertaken to collect the data (Koroluk, 1998). The advantage of a survey is that the exact data can be collected when required for a specific model. Researchers may also be able to make use of current, on-going surveys that provide environmental data. Surveys, however, can be expensive and funds may not always be available. Also, it may not be possible or efficient to collect certain types of information through a survey.
Administrative data: In many instances, data related to the environment are collected for administrative purposes, such as to administer programs or to monitor regulations. Such data could include information on fertilizer use, pesticide sales or water quality. The advantages of this type of data are (generally) low cost and availability. The disadvantage is the data may not be in the specific form needed for model development.
Remote sensing: Information from remote sensing, in combination with powerful Geographic Information Systems (GIS) that can manipulate large amounts of spatial data, is

another potential source of environmental data that could be used in model development. This source could provide population information on rotations, crop yields and production practices, reducing our reliance on costly surveys. A significant investment is going to be required before modelers and policy analysts will be able to exploit the potential that exists with remote sensing.

Research data: In most countries, public and private research are being conducted in a number of areas related to the environment. Data from various research projects are required to support the development of environmental models. For public research, the data are readily available. However, data from private research may not be publicly available. Scale will continue to be an issue with this type of data until results can be generalized over the many conditions present in the landscape at all hierarchies.

Data challenges: A simple statement about data challenges is that the policy process would like data and information on everything. This is probably not a good strategy to adopt, feasibility notwithstanding. It is not good from the perspective of privacy or from the perspective of actually using only a limited set of what could be collected at this time. This latter perspective is true even if the cost, the response burden and co-operation were not serious issues. Analysts involved with the Agriculture and Agri-Food Climate Change Table (2000) that undertook an exhaustive analysis of mitigation options for greenhouse gas emission reduction identified five information gaps (MacDonald, 2001):

- Farm management practices—fertilizer and nutrient use, land management practices, feeding strategies and manure handling essential for assessing risks
- Farm level economic data—necessary for understanding the behavioral aspects of farm decision-making through the cost structures of farming systems
- Soils data—needed for the linkage with crop and climate data to create and to validate models of cropping system adaptation, nitrous oxide and carbon dioxide emissions and the potential for carbon sequestration.
- Climate data—required for the linkage with crop yields and soil information to establish and to validate models and to extrapolate results
- Land cover—high resolution, spatially accurate determination of land cover (land use type) for all areas of the country needed to link to soils and climatic information for assessments of adaptation needs, potential and impacts. In reality these data provide specific information about farm management decisions related to land use.

Soils and land resource characteristics are relatively static and climate/weather characteristics are uncontrollable and variable within a range defined by historical records. Only decisions about management practices can be controlled to bring about changes in greenhouse gas emissions or other environmental outcomes from agriculture. What governments need to do is develop a longer term strategy based on our current understandings, then start the process of obtaining the required data. The above list is a good starting point for physical and behavioural data that are missing or incomplete in meeting today's demands.

A critical aspect is not just collecting the data, but also developing the knowledge systems in which to house all the required information. Researchers and analysts in different disciplines and research centres, both public and private, could then draw on these knowledge systems in a coordinated and coherent fashion. Many disciplines may need to draw on the data to develop their own components of an integrated model while recognizing the need to bring together the components that need to be integrated at some point and overcoming the linkage, averaging, spatial and temporal data issues. This action does not mean creating one huge model, but rather using concepts such as meta-modelling (Bouzaher et al., 1993, 1995) to develop modules that can interact to provide predictive capacity for scenario analysis purposes, which will allow alternative outcomes to be compared.

Three new features are emerging that will need to be addressed in the next generation of data systems. One is the spatial dimension discussed above and the need to have a coherent information system in which the data resides. The system must be able to combine data on an appropriate regional basis and to address the social, environmental and economic aspects of the issues. Within this context it is important to remember that although it may seem simple to ensure that the data have a spatial dimension attached, it is as critical to have the capacity to undertake analysis at the appropriate scale level for each dimension. Rarely will the scale levels coincide.

Second, many non-market effects will need to be accounted for and in many cases value will need to be used to provide a common point of reference. Data collection then moves away from what might be objectively observed into the area of subjective valuation, which is intrinsically difficult. If the process is going to function, the possible trade-offs must be evaluated in a way that is consistent with how individuals might arrive at the same decision.

The third feature is inherent in increasing economies of scale, integration and globalization. Often the market outcomes we see (i.e. prices, trade, etc.) have very little to do with the data that can be observed and collected. As the market grows more complex, for example as we move from selling commodities to selling products with specific attributes valued by the market, much of the historic information will be of questionable value due do significant structural change. The new data we require that underlies firm behaviour will be difficult, if not impossible, to obtain as concerns increase about the value of that information to competitors. Information will take on many more aspects of a private good versus a public good. At the same time, the number of attributes on which it is necessary to collect data in order to understand what is happening in the market will increase.

These three features imply a wholesale re-orientation of our data systems to ensure they can serve us well in the future. At this time, we do not see evidence of this type of fundamental change occurring in national data gathering agencies. One problem that governments are going to have to deal with is the increasing private good nature that information is taking on, and therefore less alignment may be expected in the future between public institutions seeking information and those that possess the information. Although introduced only lately, cost recovery policies may be providing the wrong incentives and sending the wrong signals to the industry in these changing times. If information has real value, why would any profit maximizer give it away? For public policy development, the question of how to collect the necessary data is of fundamental importance and governments are going to have to deal with this question in the near future. As the information value of readily available data declines, replacements must be found—replacements that are based on a much more complex reality than in the past.

4. Importance of New Information Systems to the Policy Process

Governments realize that few of the issues can be dealt with on a sectoral basis in isolation. The horizontal and vertical nature of almost any key policy issue means that decisions must be taken within a "global" context. Many indications exist of the recognition of this fact, as noted by the many impact assessments that must be undertaken to develop new laws and regulations. As the process of governing increases in complexity, the infrastructure that supports that process must also evolve. This evolution has been difficult to achieve in a period where de-regulation and less government were key policy drivers.

The allure of economic instruments highlights this predicament well. There are many good economic reasons for placing greater reliance on economic instruments where the market place will be primarily responsible for providing the necessary signals to alter behaviour. The concept is that given the right economic signals, behaviour will adjust to determine the most efficient way in which to take advantage of that signal (or keep the cost to a minimum). We will not argue against this point. The problem is that economic signals tend to be transmitted through very messy processes and there is a high probability that they will not be transmitted through the system in the simplistic manner estimated by our simple analytical tools and frameworks. Regulation has redeeming features in that it is usually highly targeted, easy to observe and corrective action is usually not difficult. None of these features exist with most economic instruments. This fact is self-evident if one looks at most income support policies that countries now use.

Targeting must be considered. To improve the efficacy of government intervention, it is necessary to impact or to change the behaviour of the identified audience. Six questions arise as to how governments determine the identified audience:

- who is in the target group
- what is their exact situation and how does this differ from some norm
- how do they interact in their local economy
- what types of alternative choices are open to them, and
- does their location make a difference.

Two immediate concerns are how to deal with the fundamental issue of confidentiality and how to understand the situation faced by those at whom the policy is targeted. Once again emerging market developments are leading us in the direction of understanding the individual as the movement away from commodity production and toward finished product production continues. This direction is in fact being encouraged by governments as a way in which the sector can improve farm and sector returns by adding greater value throughout the system.

We need to change our way of thinking from developing data bases to creating information systems that will support the development of knowledge systems needed by the policy development process. We need to develop coherent systems that will have four attributes:

- contain the data required to undertake analysis of the complexity noted above
- be constructed in such a manner as to be able to facilitate analysis at whatever spatial level is deemed appropriate for the question at hand
- possess the capacity to scale up to whatever hierarchical level is necessary, and
- provide information at all intermediate levels demanded by the different stakeholders involved in the decision making process.

For most problems dealing with agriculture, it is probably easiest to think of a system that can take us from the field level (often the lowest common denominator of decision making) up to the higher ecological areas (ecozones) and administrative regions (provinces or nations).

In the past, a substantially ad hoc process of determining what data might be required served us relatively well, as long as fairly aggregate analysis (regional or national) could be undertaken and possible economic outcomes determined. However, this analysis was done with little information on exactly what the incidence of resulting actions would be. Field level impacts were largely the domain of the physical scientists and at the farm level there was some integration of effort by economists and physical scientists. However at this scale, the information could provide the policy process with some anecdotal “representative farm” information, but not the systematic information that would be useful to the decision making process.

As an example of the progression from a data system to information system, we refer to Canada’s Whole Farm Data Base (WFDB) (Statistics Canada). The WFDB is a cooperative effort between Agriculture and Agri-Food Canada and Statistics Canada. It evolved as a response to user demands for disaggregated farm level data. The objective of the WFDB was to integrate farm level data from a variety of different surveys and sources into a single, consistent and easy to access national system. Statistics Canada hopes to eliminate the duplication of surveys, to reduce the respondent burden and to maximize the use of existing data sources. The resulting data product provides integrated income statement, balance sheet, off-farm income, crop area and livestock inventory estimates by farm type, farm size and sub-provincial region.

The WFDB, which consists primarily of financial and production data, represents the evolution from data base to information system. What is required to create a complete knowledge system is the integration of non-financial data, such as the agronomic and environmental variables, and farm and regional level modelling capacity into the system. The WFDB would then become invaluable in addressing most of the needs of government policy analysts and industry decision makers.

5. Future Direction

What this discussion suggests is that a much more co-ordinated effort will be required in the future, a direction that our current institutional structures may not serve well. The policy development process is breaking down the barriers that existed previously and allowed us to segment neatly the responsibilities for collecting and distributing different types of data from those who used the data. As the demand for analysis becomes much more horizontal and spatial in nature, so too must the systems that serve this demand change, as will the economic and social structure we are trying to affect.

Let us take the situation in Canada as a case study into the structural problems that can be encountered. The primary responsibility for collecting and distributing data lies with Statistics Canada, which collects a wide range of economic and demographic information. Statistics Canada reports to the Minister of Industry and operates under its own Act which in part ensures the confidentiality of the information it collects. It has only limited responsibility for analysis and none for policy development. AAFC however, has several branches (Strategic Policy, Market Information and Services (MISB), Prairie Farm Rehabilitation Administration (PFRA)) where the mandate for agricultural policy development resides and where most of the economic analytic capacity at the federal level resides. Within AAFC also resides the Research Branch which has a mandate to collect bio-physical information on the resource base used by agriculture and to carry out a largely physically based research program. However, the role of the Research Branch in policy development is weak and the linkages of its databases to those of Statistics Canada are also limited, generally only to Census data.

Within this system exist many components necessary to build the information and knowledge system needed for the modern policy development process at the federal level. However, at this time neither a plan nor a mandate exists to create such a system, in part due to the historic institutional divisions that are present. Statistics Canada has a mandate that covers all sectors and as such has designed its systems to meet its own requirements and mandates (i.e. production of national accounts). It is branching out into environmental or "green" accounting, but is encountering many of the same difficulties discussed above.

A critical question arises as to the ability of this current institutional structure to adapt to the evolving demands where greater co-ordination and decision making within the information system is required. Whether a wholesale redesign is required is a question that will not be answered here because there is insufficient information and analysis to determine if marginal or fundamental change is required. Nor will we recommend exactly what that evolving information and knowledge system should look like. Let it suffice to say that it will be different. There will have to be a high degree of convergence between the bio-physical, economic and demographic databases that exist now and ways of moving easily between different hierarchical levels need to be developed.

Most governments have given indications of the need to move in this direction, but change is difficult. It is hard to let go of what has worked well in the past and experiment with new ways of doing business to meet emerging demands. This is particularly true when future information demands are difficult to predict. Often the original system has to almost collapse before change can occur and this collapse may be closer than most think. For example a fundamental element of existing data is based on market prices. However, the data currently available publicly is becoming more and more irrelevant as vertical integration and contracting of production and services grow. Also, as concentration increases in some commodity sectors, more and more data will become subject to confidentiality rules and will no longer be available. Another good example is climate change, where the analytic support system for the policy development process quickly exhausted its capacity to answer the real questions being asked. The inabilities of the scientific and analytic community to supply the types of information requested are evident. Easy solutions do not exist if "science-based" responses are going to lead the way to addressing the complex issues and problems that countries confront when dealing with something as overwhelming as climate change.

The challenge is clear. Shortcomings in data are not the fundamental issues, only symptoms. The information system on which the development of a knowledge system depends must be developed. Currently we have data systems that in certain restricted ways served us well in the past and provided us with key building blocks. However, we now need to move beyond these basic building blocks to the next generation. Fortunately, two key elements are currently present—an unprecedented demand for analysis related to the policy objectives of sustainable development and a well educated and informed public that will not be satisfied with anything but the "best." Analysis needs to be based on science and information systems that underpin a knowledge system that can produce acceptable estimates of future outcomes on which to base decisions today. For analysts trying to develop models related to sustainable development, the challenge for data suppliers is a systems issues, even if it is not normally posed in this manner. This challenge is far more interesting and fundamental to addressing societal demands.

References

Agriculture and Agri-Food Climate Change Table. 2000. "Options Report: Reducing Greenhouse Gas Emissions from Canadian Agriculture." Publication no. 2028/E National Climate Change Secretariat. Ottawa, Canada.

Antle, J.M., and R.E. Just. 1992. "Conceptual and Empirical Foundations for Agricultural-Environmental Policy Analysis." *Journal of Environmental Quality* 21(3): 307–316.

Bouzaher, A., P.G. Lakshminarayan, R. Cabe, A. Carriguiry, P.W. Gassman, and J.P. Shogren. 1993. "Metamodels and Non-point Pollution Policy in Agriculture." *Water Resources Research* 29(6): 1579–1587.

Bouzaher, A., J.P. Shogren, D. Holtkamp, P.W. Gassman, D. Archer, P.G. Lakshminarayan, A. Carriguiry, R. Reese, D, Kakani, W.H. Furtan, R.C. Izaurralde, and J. Kiniry. 1995. "Agricultural Policies and Soil Degardation in Western Canada: An Agro-Ecological Economic Assessment." Technical Report, Policy Branch. Agriculture and Agri-Food Canada, Ottawa, Canada.

Dumanski, J., W.W. Pettapiece, and R.J. MacGregor. 1998. "Relevance of Scale Dependent Approaches for Integrating Biophysical and Socio-Economic Information and

Development of Agroecological Indicators." *Nutrient Cycling in Agroecosystems* 50: 13–22.

Lee, J.G., and S.B. Lovejoy. 1991. "Integrated Assessment of Environmental Effects from Agricultural Production." *North-Eastern Journal of Agriculture and Resource Economics* (April): 61–67.

Koroluk, R. (Project Leader). 1998. "Manure, Fertilizer and Pesticide Management in Canada: Results of the 1995 Farm Inputs Management Survey (FIMS)." Publication no. 1950E. Policy Branch, Agriculture and Agri-Food Canada, Ottawa, Canada.

Kulshreshtha, S.N., M. Boehm, M. Bonneau, R.J. MacGregor, and J. Giraldez. 1998. "Contributions of Canadian Agriculture to Greenhouse Gas Emissions: Preliminary Results of Selected Policy Options." *World Resources Review* 10(4): 515–584.

MacDonald, K.B. 2001. "Modelling Greenhouse Bases Emissions from the Agriculture and Agri-Food Sector: Assessment of Data Gaps." Report prepared for Strategic Policy Branch, Agriculture and Agri-Food Canada, Ottawa, Canada.

Statistics Canada. 2001. "Whole Farm Data Base Reference Manual." Catalogue no. 21F0005GPE, Ottawa, Canada.

The Use of Agri-Environmental Indicators for Policy Analysis and Projections: The OECD Experience

Kevin Parris[1]
Policies and Environment Division, Directorate for Food, Agriculture and Fisheries,
OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.
e-mail: kevin.parris@oecd.org

Abstract: This paper (Section 1) examines what are the policy issues which agri-environmental indicators need to address. Measuring environmental impacts (Section 2) is discussed in terms of agriculture's role in protecting the stock of natural resources, reducing environmental pollution, and improving farm management and resource use efficiency. The paper concludes by examining (Section 3) future directions indicator work may take for policy purposes.

Keywords: Agri-environmental indicators, policy analysis, projections.

1. What are the Policy Issues Which Indicators Need to Address?

Globally the scale of agricultural expansion is going to place greater pressure on the environment over the coming decades if it is to meet the needs of an additional 1.5 billion in the global population expected by 2020. As well as producing food and fibre, agriculture is also increasingly being required to provide various environmental goods and services, such as serving as habitat for wildlife; providing ecological functions, for example, acting as greenhouse gas sink; and supplying amenities, like landscapes.

Some view agriculture as entering a new era of expansion through advances in biotechnology and the revolution in information and communication technologies. The process of trade liberalisation and globalisation of the agro-food chain may also provide the basis for the investment and future growth of agriculture on an environmentally sustainable path. But others are more pessimistic and see signs that current farming practices are leading to the degradation and depletion of the natural resource base upon which farming depends, namely soils, water, natural plant and animal resources.

Changes in policy settings are a key influence on the environmental performance of agriculture. With the greater public and policy focus on agri-environmental issues emerging over the 1990s, this trend is likely to continue with implications for agriculture and the environment because governments can be expected to:

- reform agricultural policies, especially the level and composition of support;
- further develop agri-environmental measures; and,

[1] The author wishes to thank OECD colleagues for their help in preparing this paper. Any remaining errors in the paper are the responsibility of the author, and the views expressed do not necessarily reflect those of the OECD or its Member countries. For further information on the OECD work on agri-environmental indicators please contact Kevin Parris at the address above or visit the OECD website at: http://www.oecd.org/agr/env/indicators.htm

- strengthen environmental policies both domestically and multilaterally.

The reform of agricultural policies, by reducing the overall level and composition of agricultural support (at US$327 billion for OECD countries in 2000, see OECD, 2001a), should improve the domestic and international allocation of resources, reduce incentives to use polluting chemical inputs and to farm fragile land. Such reforms will tend to reverse the harmful environmental impacts associated with commodity and input specific policy measures. In those cases where environmental effects, both harmful and beneficial, are not taken into account by farmers policy reform may not improve environmental quality, therefore, targeted environmental measures might also be needed (OECD, 2001b).

As part of the agricultural policy reform process many OECD countries started to introduce ***agri-environmental measures,*** beginning around the late 1980s. While the nature of these measures varies across countries, they have mainly focused on altering those farm management practices, sometimes encouraged by high price support levels, incompatible with achieving environmental objectives. These measures include the provision of payments, such as those for the adoption of low-input farming systems, which at present represent under 10% of total OECD agricultural support.

A few countries have also used taxes to limit the pollution from the use of pesticides and fertilisers, and enforced restrictions on farmers to meet certain minimum standards, such as the disposal of animal waste into watercourses. Also, land diversion schemes, although in most cases originally introduced to achieve supply control objectives, are increasingly including environmental conditions, such as diverting land for habitat use to encourage wildlife and to help reduce soil erosion. A number of countries also use voluntary efforts, including farm advisory services, to address local and community related environmental issues, and raising environmental awareness amongst farmers.

While evidence is still limited, the introduction of agri-environmental measures have helped to alter farm management practices and change land use patterns. These changes have contributed to, for example, the conservation of wildlife habitats and the reduction of diffuse pollution. But there is at present insufficient information in many cases to be sure about the extent and permanence of these changes across OECD countries. In some cases improvements have been made, but have been more costly than would have been the case in the absence of production enhancing policies. Also, adverse environmental impacts still remain at relatively high and damaging levels in many cases.

Future ***domestic environmental measures and multilateral environmental agreements*** may also have a greater influence on agriculture mainly because:

1. Progress in reducing ***environmental pollution*** from industrial and household waste has shifted the focus to agriculture as the share of agriculture in total emission loadings for certain pollutants, especially nitrates and phosphates, has been rising.
2. Given that agriculture is the major user of land and water for most OECD countries, environmental policies that address ***resource depletion issues***, and biodiversity, habitat and landscape concerns, inevitably involve agriculture.
3. There are an increasing number of ***multilateral environmental agreements*** with implications for agriculture, some operating regionally (e.g. European Landscape Convention), and others globally (e.g. the Convention on Biological Diversity.

3. Measuring the Environmental Impacts of Agriculture

Agriculture's impact on the environment involves a sequence of processes. The quantity of agricultural production is affected by the financial resources available to agriculture (both returns from the market and government support), the incentives and disincentives facing farming, and the kinds of management practices and technologies adopted by farmers. These practices and technologies impact on the productivity of the natural resources (e.g. soil) and purchased inputs (e.g. fertilisers) used by farmers. Depending on the management and productivity of agriculture's use of resources and inputs this will affect the rate of depletion and degradation of soils and water, the flows of harmful emissions (e.g. nutrients) into soils, water, air and the atmosphere, and the quantity and quality of biodiversity, habitats and landscape features.

OECD is developing a set of agri-environmental indicators to measure the environmental performance of agriculture, taking into account the:

- policy challenges which indicators need to address;
- sequence of processes that describe how agriculture impacts the environment; and
- experience of OECD countries and other international efforts to develop indicators.

This section examines the use of the OECD indicator set to measure the environmental impacts of OECD agriculture since the mid-1980s, in terms of agriculture's role in: [2]

- Protecting the stock of natural resources and landscapes impacted by agriculture.
- Reducing environmental pollution from agriculture.
- Improving farm management practices and resource use efficiency.

2.1. Protecting the Stock of Natural Resources and Landscapes Impacted by Agriculture

Agriculture plays a critical role in the conservation (or depletion) of the stock of natural resources used for production, notably soil and water resources, because for most countries agriculture accounts for the major share in the use of these resources. Farming activities also impact positively and negatively on the quality and quantity of natural plant and animal resources and landscapes.

Agriculture accounts for 40% of total OECD ***land use***, with about 3% of this expected to be converted to other uses by 2020, mainly the reversion of 'marginal' farm land to forests or 'natural' habitats, such as reclaimed wetlands (OECD, 2001c). Further urban encroachment is also expected to lead to the loss of some highly productive agricultural land, a development which is in general irreversible.

Poor ***soil conservation*** practices on agricultural land can increase rates of water and wind erosion above those levels that occur naturally. While the area of agricultural land at high or severe risk to erosion is not extensive across OECD countries (i.e. above 22 tonnes/hectare/year, compared with rates of 1–5 tonnes/hectare/year, depending on soil conditions, considered as tolerable rates of erosion), for some countries more than 10%

---

[2] This section draws, in particular, from the OECD work on agri-environmental indicators (see the report OECD, 2001b, which also provides detailed descriptions of the method used to calculate these indicators) and the use of the indicators in projections of the environmental performance of OECD agriculture to 2020 (see OECD, 2001c; and 2001d).

of agricultural land falls within this risk class. Trends in soil erosion for OECD countries, notably Australia, Canada and the United States, show a reduction from high/moderate into tolerable/low classes of erosion over the 1990s. This has mainly resulted from the effects of the adoption of conservation tillage practices, less intensive crop production and the removal of some marginal land from production.

Despite the recent improvements in soil conservation practices, estimates of ***the economic costs associated with soil erosion*** are significant. In Australia, for example, it is estimated (Industry Commission, 1996) that the value of agricultural production foregone as a result of soil degradation was between 5% to 6% of the total annual value of production in 1994-95, while annual damage from off-farm soil sedimentation of rivers, lakes and reservoirs was estimated in the United States to be between USD 2–8 billion in the mid-1990s (USDA, 1997).

Even with greatly improved irrigation ***water use*** efficiency, it is anticipated that one-third of the world's population will remain short of water by 2025 (Seckler, et al, 1998). In OECD countries agriculture currently accounts for 45% of total water utilisation, and for over 60% in nearly a third of OECD countries. Trends in agricultural water use are mainly affected by the market prices for irrigated crops and related changes in the area of irrigated crops, the level of subsidy provided to farmers for water use, agricultural support overall, and improvements in water use efficiency.

For most OECD countries ***water charges*** for farmers are substantially below those paid by other users, due in part to government support for irrigation infrastructure and water delivery costs, as well as the type of water rights, and the pricing criteria and type of charges levied by water suppliers. Some caution, however, is required in drawing comparisons between water prices paid by the different users, because water supplied to agriculture is usually of lower quality than that used by households. Also, the capital and running costs of the water conveyance system is generally lower for agriculture than for households or industry. Many OECD countries, however, are beginning to move toward full cost recovery for water supplied to agriculture, although it is too early in most cases to know the impacts of these reforms.

There is increasing policy focus on the relationship between ***agriculture, biodiversity and habitats.*** Selective plant and animal breeding programmes drawing on ***genetic material*** from on-farm and genebank sources, have helped to raise agricultural productivity. In the US, for example, it is estimated that over the past 60 years, half of agriculture's productivity increases can be attributed to genetic improvements. The extent to which these improvements can be maintained will be influenced by the diversity of crop varieties and livestock breeds used in production; the extent of the loss of genetic resources that could help to raise productivity; the impact of biotechnology; and government policy which may either limit or promote certain biotechnologies according to their food safety and environmental impacts.

The impact of agriculture on the quality and quantity of ***species and eco-system diversity*** is to a large extent determined by the expansion (or contraction) of the farmed land area, and the intensity of agricultural production in terms of input use and farming practices. A number of agro-ecosystems can serve to maintain wild species diversity, such as some pasture and grassland systems. Over recent decades the intensification of agricultural production and changing farm practices (e.g. crop rotations) have led to harmful impacts on wild species and the destruction of wildlife habitats. Also, in some marginal farming areas agricultural land has been converted to other land uses, especially forestry, with the loss of certain agro-ecosystems.

The adverse impacts of ***non-native species*** on agriculture and agro-ecosystems is perceived as one of the most important issues in natural resource management in some OECD countries. Economic losses from non-native species in the US, for example, were estimated over the 20th century at USD 97 billion (OTA, 1993). Some researchers consider non-native species will continue to place a major burden on farming and the environment, by inflicting financial losses to farmers through damage to crops and competition for livestock forage, and through predation leading to the destruction and decline of native species. In some cases these non-native species can be beneficial helping to increase food production or by acting as biological control agents.

Regarding changes in ***agricultural landscapes*** there does seem to have been a trend towards increasing homogenisation of landscape structures in OECD countries over the past 50 years, including the loss of some cultural features (e.g. stone walls). This trend is closely related to the structural changes and intensification of production, but since the late 1980s, the process toward increasing homogeneity of landscapes could be slowing or in reverse in some regions. To establish the value the public places on landscape non–market valuation techniques can be used. Studies in those countries using these techniques show that agricultural landscapes are highly valued in many cases, but there is a large variation in the values estimated (OECD, 2001b).

2.2. Reducing environmental pollution from agriculture

Flows of materials into water (e.g.nutrients, pesticides) and emissions into the air (e.g. ammonia, greenhouse gases) are an inevitable part of agricultural production systems. Reducing the flows of these materials and emissions to an ‘acceptable’ level of risk in terms of human and environmental health is a priority for policy makers.

A key issue regarding agriculture and ***water quality*** relates to nitrate pollution in surface and groundwater and phosphorus levels in surface water. An excessive level of agricultural nutrients in water is both a human health concern, since it impairs drinking water quality and an environmental burden, since it can cause eutrophication (i.e. algae growth in water). Agriculture is the major source of nitrates and phosphates polluting aquatic environments in most OECD countries, accounting for more than 40% of all sources of nitrogen emissions and over 30% of phosphorus emissions into surface water. The extent of groundwater pollution from agriculture is less well documented, partly because it can take many years for nutrients to leach through soils into aquifers.

Given the reduction in ***nitrogen surpluses*** since the mid-1980s for most OECD countries, agricultural nutrient pollution of water is probably declining. But there are areas where the overall levels are extremely high. And for those countries where livestock production has expanded, such as Canada, Ireland, Korea, New Zealand and the United States, nitrogen surpluses have been growing. Many countries are addressing these problems by encouraging management practices that improve farm input use efficiency, and in some cases reducing animal numbers, such as in The Netherlands.

Similarly the reduction in ***pesticide use*** in many countries, would also suggest that water pollution from this source may be declining, although for some countries pesticide use has risen in response to an expansion in crop production. Even so, the long lag between pesticide use and their detection in groundwater means that, as with nitrates, the situation could deteriorate before it starts to improve. Estimates of the costs of water pollution, indicate that lower pollution levels could bring significant benefits. In the United Kingdom, for example, the annual external cost of agricultural water pollution in

1996 was estimated at GBP 231 million (USD 360 million), of which just over half was the cost of removing pesticides from drinking water (Pretty, et al 2000).

Pesticides can also pose health risks to those applying pesticides in the field and in close proximity to land treated with pesticides. But the incidence of these health impacts is poorly documented for most OECD countries. Where information is available the extent of the problem appears limited, and could decline further with improvements in application technologies and practices, education and training. Levels of pesticide residues in foodstuffs for most countries are below current maximum permissible levels, although on occasions these limits are exceeded for fruit and vegetables.

The impact of pesticides on wildlife is also poorly reported in most OECD countries, but where adverse effects occur these could diminish as farmers substitute more broad scale for narrow spectrum pesticides, and increasingly use precision farming technologies to apply pesticides. This should help to avoid harmful impacts on beneficial wildlife, such as pollinators, and non-target flora and fauna. Also important in this context are the scientific uncertainties related to the possible risks associated with ***pesticide use and endocrine disruption***, with recent research showing that the chemicals in some pesticides, are disrupting human and wildlife endocrine systems (hormonal systems), with harmful impacts on human and wild species fertility and pregnancy. This problem has provoked the US and the European Union, for example, to recently introduce endocrine disrupter screening programmes.

One of the main ***air pollutants*** from agricultural activities is ***ammonia*** (NH_3), which can lead to plant foliage damage, soil acidification and eutrophication. Evidence for some European countries indicates that around 95% of ammonia emissions into the air result from agricultural activity, with about 60% from animal manure and much of the remainder from the use of chemical nitrogen fertilisers (UNEP, 1999).

With increased atmospheric concentration of greenhouse gases (GHGs) contributing to the process of ***climate change*** and global warming, most OECD countries, under the 1994 United Nations Framework Convention on Climate Change, have committed themselves to stabilise GHG emissions at 1990 levels by 2000. Most OECD countries have also agreed to implement the 1997 Kyoto Protocol, which specified the levels of total national emissions for the target period 2008 to 2012.

Agricultural GHG emissions contribute about 8% of total OECD emissions (in CO_2 equivalents). While the contribution of OECD agriculture to the main GHG gas, carbon dioxide is only about 1%, it accounts for 60% of nitrous oxide, and nearly 40% of methane emissions. Livestock enteric fermentation, manure and fertiliser use account for most agricultural GHGs, but the shares of other emission sources are also important in some cases, notably crop residues, biomass burning, and wetland rice cultivation.

Currently there are no systematic estimates across OECD countries of the ***capacity of agriculture in sequestering (removing) carbon in soils***. The carbon sequestration capacity of agriculture is affected by a complex set of relationships, but estimates show that about 50% can be achieved by adopting soil conservation and improving crop residue management (e.g. reduction of stubble burning), 25% by changing cropping practices (e.g. increases in soil cover), and much of the rest through a combination of land restoration efforts and converting cropland to pasture (Antle, et al 2000). Future changes in sequestering carbon by altering farming practices and production intensity, is thought to increase soil carbon slowly over the first 2 to 5 years, with larger increases between 5-10 years, reaching a finite limit after about 50 years.

Agriculture also has the potential to reduce GHG emissions through the replacement of fossil fuels with ***biomass energy***, from crops. International Energy Agency (IEA) projections expect non-hydro renewable energy (NHRE) sources (mainly geothermal, solar, wind, tide and biomass) to be the world's fastest growing primary energy source up to 2020 at nearly 3% per annum (IEA, 2000). Most of this is accounted for by OECD countries, and the contribution of biomass in world total NHRE may decline from nearly 75% to about 50%. However, the share of NHRE energy sources in total OECD electricity production is small but projected to rise from 2%–4% between 1997 to 2020.

2.3. Improving farm management and resource use efficiency

The projected decrease in international agricultural commodity prices, in real terms by 2020 (FAO, 2000; OECD, 2001c), can be expected to bring continued pressure on farm incomes and contribute toward ***structural changes in OECD agriculture***, leading to a further reduction in the share of agriculture in GDP and total employment. These developments suggest average farm size will continue to increase in terms of area and capital assets for most countries in a move towards further gains in productivity to support agricultural profitability. Major drivers in agricultural profitability and structural change, apart from changes in market conditions, are developments in technologies, farm management practices, and resource use efficiency.

Many of the ***technologies*** available to farmers have the potential to steer agriculture along a sustainable path, providing both economic and environmental benefits. Examples include, precision farming, such as linking global positioning to geographical information systems to map precise fertiliser and pesticide requirements; and biotechnology, for example, genetically modified (GM) crops that are insect and herbicide resistant.

Adoption of environmentally benign management practices by farmers is also heavily dependent on profitability, risk perceptions, and the extent to which the regulatory system restricts the use of certain farming practices. Even where innovative management practices are profitable, there can be impediments affecting their rate of adoption, such as skills of farmers and different perceptions of economic risks. Illustrative is the small number of OECD countries where more than 40% of farmers have even basic post-school training.

Improving the ***efficiency of using farm inputs*** (i.e. energy, fertilisers, pesticides, water) is important in terms of reducing potential pressure on the environment and also lowering rates of natural resource depletion. Farm input use efficiency is affected by resource prices (costs relative to farm receipts), the availability and cost of technologies which can improve efficiency, and the effects of government policies on input use. In the United States, for example, maize production per kilogram (kg) of nitrogen applied rose from 18 to 25 kg between 1985 to 1995 (Isherwood, 2000). But there can be trade-offs in changing to more environmentally benign farming practices, where for example, the switch from conventional to conservation tillage to combat soil erosion has involved higher uses of herbicides to control weeds

While the continued use of inputs and the adoption of new technologies and management practices by farmers will be necessary if agriculture is to achieve further improvements in productivity, there are uncertainties about the ***limits to agricultural productivity gains imposed by physical and biological environmental constraints***. Technological improvements and increased input use might be unable to raise agricultural production sufficiently to offset the depletion of soil and water resources

(Brown, 2000). It is also thought that in some regions further intensification of agriculture can induce irreversible changes in ecosystems once sustainable thresholds of natural ecosystems are exceeded, especially soil degradation and depletion of water resources. Other examples of biophysical constraints include loss of agricultural genetic resources, and pest and disease resistance to pesticides.

These concerns, together with other related issues such as food safety and quality, are fuelling a discussion about the what to which the ***future development of organic agriculture*** could help overcome these biophysical environmental constraints. While organic farming has grown rapidly over the 1990s, nevertheless, its share of the OECD total agricultural area is under 2%. The future expansion of organic farming will largely depend on policy incentives, raising yields, lowering producer conversion costs, and reducing consumer prices.

With the current yields of organic farming, an expansion of organic farming would involve both an increase in the area cultivated and animal stocking rates if current production levels were to be maintained. This could conflict with the conservation of biodiversity and habitats if additional 'high nature value' land were brought into production. However, in comparing the yields, costs and prices of conventional versus organic farming, no account is usually taken of the relative environmental costs associated with the two systems in terms of the effects on soil degradation, water depletion and pollution, and effects on human health and wildlife.

3. Future Directions for Work on Indicators

For some agri-environmental issues there is incomplete knowledge and data to establish trends. Information is incomplete, for example, concerning the degree of groundwater pollution or rate of depletion resulting from farming, and the human health and environmental risks associated with pesticide use. In other cases the linkages between different indicators are understood but are not easy to measure, such as between changes in farm management practices and environmental outcomes, or attributing the relative impact of agriculture and other activities, for example, on water pollution. Also for a number of areas, notably agriculture's impact on biodiversity, habitats and landscape, the understanding and measurement of these impacts is still at a preliminary stage of research. This is partly because of the high costs associated with monitoring programmes.

The overall direction for further developing agri-environmental indicators is to meet the objectives of providing information on the state and changes in environmental conditions in agriculture, and using indicators for policy monitoring, evaluation, and projections. To meet the objectives for an improved agri-environmental information base for policy makers and the wider public will require (OECD, 2001b):

- improving the analytical soundness and measurability of indicators, such as a better understanding and measurement of agriculture soil carbon sinks, and also how to best track agriculture's impacts on biodiversity, habitats, and landscapes;

- overcoming data deficiencies and improving interpretation of indicator trends, especially through better expression of the spatial variation of national level indicators, and developing appropriate baselines, threshold levels and targets to help assess policy performance;

- using agri-environmental indicators to better inform policy monitoring, evaluation and projections, for example, determining the effects of irrigation water and infrastructure support on irrigation management and water use; and,

- developing indicators that can help to examine synergies and trade-offs between the economic, social and environmental dimensions of sustainable agriculture.

References

Antle, J.M., S.M. Capalbo, S.Mooney, E.Elliot, and K.Paustian (1999), *Economics of Agricultural Soil Carbon Sequestration in the Northern Plains*, Research Discussion Paper No.38, (revised March 2000), Montana State University-Bozeman, Montana, United States. Available at: http://agecon.lib.umn.edu/cgi-bin/isearch

Brown, L.R. (2000), "Challenges of the New Century", Chapter 1, in Worldwatch Institute Report, *State of the World 2000*, Washington, D.C., United States.

FAO (2000), *Agriculture: Towards 2015/30 - Technical Interim Report*, Rome, Italy.

IEA [International Energy Agency] (2000), *World Energy Outlook 2000*, see p.24 and pp.102- 103, IEA, Paris, France.

Industry Commission (1996), *Land Degradation and the Australian Agricultural Industry*, Staff Information Paper, Australian Government Publishing Service, Canberra, Australia.

Isherwood, K.F. (2000), *The State of the Fertiliser Industry Past, Present and Future*, paper presented to the 68th International Fertiliser Association Annual (IFA) Conference, Oslo Norway. Available at the IFA web site: http://www.fertilizer.org/PUBLISH/oslo00.htm

OECD (2001a), *Agricultural Policies in OECD Countries - Monitoring and Evaluation 2001*, Publications Service, Paris, France.

OECD (2001b), *Environmental Indicators for Agriculture Volume 3: Methods and Results*, Publications Service, Paris, France. The *Executive Summary* of this report is available from the OECD agri-environmental indicators web site at: http://www.oecd.org/agr/env/indicators.htm (see "Publications").

OECD (2001c), *OECD Agricultural Outlook 2001–2006*, 2001 Edition, Publications Service, Paris, France.

OTA [Office of Technology Assessment] (1993), *Harmful nonindigenous species in the United States*, United States Congress, OTA-F-565, US Government Printing Office, Washington, D.C., United States. Available at: www.ota.nap.edu/pdf/1993idx.html

Pretty, J.N., C. Brett, D. Gee, R.E. Hine, C.F. Mason, J.I.L. Morison, H. Raven, M.D. Rayment and G. van der Bijl (2000), "An assessment of the total external costs of UK agriculture," *Agricultural Systems*, Vol.65, pp.113-136.

Seckler, D., U. Amarasinghe, D. Molden, R. de Silva and R. Barker (1998), *World Water Demand and Supply, 1990 to 2025: Scenarios and Issues*, Research Report No.19, International Water Management Institute (IWMI) Colombo, Sri Lanka. Available at the IWMI web site: http://www.cgiar.org/iwmi/pubs/indexp.htm

UNEP [United Nations Environment Programme] (1999), *Global assessment of acidification and eutrophication of natural ecosystems*, UNEP Technical Report, Nairobi, Kenya. Available at the UNEP web site: http://www.grid.unep.ch/geo2000/techrpts.htm

USDA [United States Department of Agriculture] (1997), *Agricultural Resources and Environmental Indicators, 1996-97*, Agricultural Handbook No. 712, Natural Resources and Environment Division, Economic Research Service, Washington, D.C., United States. Available at: www.ers.usda.gov .

Using Survey Data and Modeling to Assist the Development of an Agri-Environmental Policy

J. Jeffery Goebel, Robert L. Kellogg
Natural Resources Conservation Service, U.S. Dep. of Agriculture
5601 Sunnyside Ave., Mail Stop 5475, Beltsville, MD 20705-5475, USA,
e-mail: jeff.goebel@usda.gov, robert.kellogg@usda.gov

Abstract: Data gathering operations are typically designed and implemented based upon requirements stipulated by the sponsors. However, the database and information derived from a scientifically based study can be just as valuable in addressing issues that were not specified when the study was designed. This is particularly true for those involved in agricultural statistics and analysis of agri-environmental issues. The National Resources Inventory (NRI) is a longitudinal survey of soil, water, and related environmental resources designed to assess conditions and trends on non-Federal U.S. lands, with the specific goal of supporting agricultural and environmental policy development and program implementation. The NRI has served as the basis for a number of analyses dealing with a variety of natural resource issue.

Keywords: Agri-environmental policy, Simulation modeling, Carbon sequestration, Soil erosion, Agricultural statistics, Survey sampling, Imputation

1. Introduction

Perspectives and policies regarding U.S. agriculture have evolved significantly over recent decades. Agricultural production has often been perceived as degrading the environment, but public programs and landowner commitments now show that agriculture can serve as a source of enhancement for the environment as well.

In the 1980's, agri-environmental policies were developed with a focus on soil conservation as a means to preserve agricultural productivity; then objectives were broadened to include, for example, water quality, wildlife habitat and air quality. Farm income support has continued to be a key objective of agri-environmental programs (USDA, 2001). The conservation provisions of the *Food Security Act of 1985* and the *Food, Agriculture, Conservation, and Trade Act of 1990* changed the priorities of Federal soil and water conservation agencies, their state and local cooperators, and the farmers. Provisions created incentives to reduce soil erosion on agricultural lands, enhance wildlife habitat, conserve wetlands, and improve water quality as means to protect and enhance the environment (Flather et al., 1992; USDA, 1999). Included were programs such as the Conservation Reserve Program, Cross Compliance, Swampbuster, and Sodbuster.

Political factors, public awareness and education, and technological advancements have all been factors in development and implementation of these policies and programs. Scientists, economists, and policy makers have found the National Resources Inventory (NRI) to be a source of scientifically credible and nationally consistent data that helps them formulate policy proposals and analyze economic and environmental impacts. The NRI had been developed in the 1970's by the United States Department of Agriculture (USDA) as a tool to assess status, condition, and trend of soil, water, and related resources on the Nation's nonfederal lands, as mandated by the Rural Development Act of 1972 and the Soil and Water Resources Conservation Act of 1977.

A variety of natural resource issues have been analyzed using the NRI. These issues include: land use change with emphasis on loss of agricultural lands to urban development; conservation provisions of the 1985 Farm Bill dealing with soil erosion and conversion of wetland ecosystems; transport of agricultural chemicals into water supplies; the role of agriculture in sequestering carbon. Each issue is briefly addressed in later sections of this paper. The use of the NRI as an analytic framework is highlighted.

2. The National Resources Inventory

The National Resources Inventory (NRI) is a longitudinal survey conducted by the USDA Natural Resources Conservation Service (NRCS), in cooperation with the Iowa State University Statistical Laboratory. The purpose of the NRI is to provide support for agricultural and environmental policy development and program implementation. The NRI is a panel survey of land use and associated natural resource attributes, conducted at 5-year intervals from 1982 through 1997.

The NRI sample is a stratified two-stage unequal-probability area sample (Nusser and Goebel,1997; Goebel and Baker, 1987). The primary sampling units (PSUs) are areas of land called segments. The segments vary in size, from 16 to 256 hectares. Sampling rates vary across strata, typically being between 2% and 6%; there are 300,000 sample segments in the current national sample. Data are collected for the entire segment for some items such as developed land and water area. Detailed data on soils, land use, and numerous conservation issues are collected at a randomized sample of points within the segments. Generally there are three points per segment, but some segments only contain one or two points. Some data are collected on a census basis external to the sample survey; these include total surface area, Federal land area, and area in large water bodies for hydrologic unit portions of counties.

A fairly complex estimation process has been developed in order to provide a final data set that has only one set of weights and is relatively simple to use (Fuller, 1999; Nusser and Goebel, 1997; Breidt et al., 1996). This data set must contain all segment and point sample data (for 1982, 1987, 1992, 1997), must incorporate the county and sub-county ancillary data as controls, and must incorporate thousands of estimates published in 1982 and 1992 as controls. The estimation process incorporates several imputation procedures, pseudo point generation, small area estimation, ratio estimation, and raking.

The NRCS and its predecessor, the Soil Conservation Service, have a long history of conducting natural resource inventories and monitoring [Goebel (1998); Nusser and Goebel (1997)]. Reconnaissance methods were used in 1934 and 1945, when there was concern about the effects of droughty conditions, inadequate land management practices, and soil erosion on agricultural production. Statistical sampling methodology was introduced in 1956, when the Soil Conservation Service was assigned responsibility to lead a cooperative endeavor of eight Federal agricultural agencies to develop estimates of the magnitude and urgency of the various conservation measures needed to maintain and improve the country's productive capacity (USDA, 1962). Further studies were conducted in 1967, 1975, and 1977 as there were needs for more up-to-date information. The 1977 inventory is considered the starting point for the current NRI program. It was the first study to use a survey sampling approach to nationally examine water and wind erosion, conservation practices, incidence of wetlands, and flooding propensity. However, between 1977 and 1982 there were significant changes in data collection protocols and in the sampling; therefore, the current longitudinal data set does not include data for 1977 and uses 1982 as the first point in time for temporal comparisons and analyses.

The objectives of NRCS resource inventories have expanded over time, as the focus of agricultural policy has moved toward a balance between short-term production goals, long-term capabilities, and environmental quality. Statistical techniques and data collection protocols have evolved as inventory goals have become broader and more sophisticated, and operational features have been modified because of agency reorganizations and the need for better cost-effectiveness. This evolution is continuing as the NRI is now moving into a "continuous inventory process" (Goebel, 1998).

3. Land Use Conversion and Loss of Prime Farmland

Analysis of agri-environmental natural resource issues requires an understanding of the dynamics of land use change, which requires consistent longitudinal data. Proper analyses also require data on how the land is being managed and on inherent natural resource parameters such as soil quality and other soil-related capabilities and characteristics.

Recent figures show that the U.S. as a whole is quite heterogeneous when it comes to land use and cover. Federal land, forest land, and rangeland each comprise about 21% of the landscape, and about 19% is cropland; the remaining area consists of pastureland, developed land, water areas, and various miscellaneous rural areas (USDA, 2000). The perspective is quite different on a regional basis – forest land predominates in the eastern half of the country, except for those areas in the center of the country where cropland is concentrated. Rangeland and Federal lands predominate in the western half of the country, which is characterized (in general) by lower rainfall, more mountainous terrain, and soils less suitable to cultivation. Even though these percentages have remained fairly stable over the past decade, there is a need to examine lands that do change cover/use over time to determine what impacts may eventually occur if these trends continue.

During the 15-year period 1982 - 1997, 11% of the nation's privately owned lands have experienced one or more major change in land cover/use. Cropland experienced the most changes – of the 168 million hectares that were cropland in 1982, almost 28 million have been converted to some other use; about 43% of these conversions were due to "land retirement" into the Conservation Reserve program; another 10% were converted into developed land and permanently lost to the rural land base. These losses and conversions were partially offset by the addition of nearly 11 million hectares of new cropland, 58% of which came from pastureland. These data on land cover/use distributions and dynamics come from the 1997 NRI. They will be used to examine many agri-environmental issues.

Several studies have examined data from the 1992 NRI to determine the effects of urbanization on both agricultural production and agri-environmental factors. The report *Farming on the Edge* (Sorensen et al., 1997) used the 1992 NRI as the framework for their examination; the NRI framework was augmented by linking in local NRCS soils data (designating certain unique and locally important soils) and agricultural production data from the Census of Agriculture. This analysis was based upon the NRI cartographic database (Kellogg et al., 1992) that consists of 33,000 spatial mapping units created by the intersections of counties, watershed boundaries, and Major Land Resource Areas. NRI data can be mapped to these polygons to enable geospatial analysis of numerous natural resource issues. The analysts established statistical thresholds to quantify "high quality farmland" and "high development". One portion of their findings was to identify the Nation's most endangered regions, using three basic factors: (1) market value of agricultural production, (2) development pressure, and (3) land quality. Identified were 20 threatened Major Land Resource Areas; they account for only 7% of the Nation's land but 21% of the Nation's prime or unique farmland lost to development, and are responsible for 51% of U.S. fruit production, 39% of vegetable production, and 28% of dairy production. Work is underway to update this study using 1997 NRI data; this could produce some changed results since the rate of conversion of rural land to development during 1992-1997 was more than 50% greater than during the period 1982-1992.

4. Potential Groundwater and Surface Water Contamination

The NRI was recently used to help study potential groundwater and surface water contamination resulting from manure nutrients (Kellogg, 2000). Manure application to agricultural lands leads to susceptibility for leaching and runoff, and hence potential contamination of water supplies. This issue is presently under discussion at scientific, program, and policy levels within the U.S.

The potential for manure nutrients to move from farm fields to groundwater and surface water is currently being examined through development of vulnerability indices. Although this issue should be addressed with hydrologic modeling at the watershed level, such procedures are not yet developed and it is necessary to use a modeling approach that simulates movements of nutrients from farm fields. A survey sampling approach is suggested, since it is not possible to examine the millions of such fields in the country.

Appropriate data do not exist for any large-scale scientific sample of fields, so an analytic framework was developed utilizing the NRI sample points. Each NRI sample point is treated as a "representative field". The sampling framework (through the sample unit weights, or expansion factors) provides the mechanism for expanding estimates up to a hydrologic unit or watershed basis. Estimates were derived for three environmental indicators of vulnerability – of movement of nutrient materials from the field. These are:

- Percolation factor - measures potential for leaching
- Annual runoff factor - measures potential for dissolved contaminants to run off a field
- Soil erosion factor - measures potential for nutrients absorbed onto soil particles to move from farm fields.

Daily climate data for 25 years for 1,473 climate stations were used to produce monthly averages, for imputation onto each NRI data point. Data on manure nutrients available for application to cropland and pastureland were converted from a county basis to a hydrologic unit basis. Vulnerability indices were derived to represent the potential for loss of manure nitrogen and phosphorus from farm fields: (1) nitrogen dissolved in runoff, (2) nitrogen in leachate, and (3) phosphorous absorbed to soil particles and removed by water erosion. Water erosion data are available for each NRI sample point in terms of the factors that go into the Universal Soil Loss Equation (USLE). Watershed vulnerability indices were derived and mapped; comparisons can be made to manure availability to examine the effects of adjusting manure loadings based upon vulnerability. These analyses help indicate the locations of priority watersheds (Kellogg, 2000); addressing concerns in these regions would have the greatest environmental impact, if certain policies are adopted. Focusing efforts in these watersheds would impact NRCS staffing and allocation of resources.

5. Soil Carbon Storage

There is currently considerable interest in the U.S. regarding soil carbon storage in agricultural cropland soils. There is particular interest at the national and international levels because carbon sequestration in soil can contribute to greenhouse gas mitigation.

There has been a desire to quantify effects of changes in agricultural practices, and to determine if farmers could receive economic benefits for helping offset (or balance) emissions of carbon dioxide (CO_2) and other greenhouse gases from non-agricultural sources. Some feel that agriculture could receive these soil carbon credits from the various industrial entities that create CO_2 emissions, for a period of up to 25 years, while technologies are developed to control present emission sources. Increased carbon storage also leads to improved soil, water, and air quality; for example, there would be decreases in risks of soil erosion and increases in soil buffering, nutrient retention, and productivity.

There are many problems associated with direct measurement of soil carbon (Campbell et al., 2001); increases are small relative to base amounts in soil, changes are slow, and distributions are quite variable. It would be impossible to carry out direct measurement of

soil carbon storage on a national basis; it is suggested that use of a simulation model is a more realistic approach (Campbell et al., 2001; Goss et al., 2001; Eve et al., 2001).
Agricultural practices such as tillage, drainage, removal, burning, and harvest have long served to induce soil carbon loss. Researchers feel that this trend has been reversed over the past 20 to 30 years, as farmers started using conservation tillage techniques and reduced use of the moldboard plow that opened up the black prairies and caused a significant soil carbon drain A number of factors have been identified that can slow depletion of soil organic matter or contribute to an increase in soil carbon storage [Lal et al (1998)]; these include:

- Cropping practices
- Additional forage crops
- Reduced tillage intensity
- Soil erosion control
- Climate change
- Proper fertilization
- Reduced summer fallow
- Chemical fallowing
- Incorporation of crop residues in soil
- Replacing annual or seasonal crops with tree crops

A modeling approach needs to account for these factors as well as soils, climate, and historical use and management of the land. The modeling must be performed in a context that accounts properly for the actual occurrences and combinations of these factors; the NRI provides an appropriate analytical framework. We briefly discuss four modeling approaches that have used or will use NRI data to make national assessments of carbon sequestration.

(1) Intergovernmental Panel on Climate Change (IPCC) Inventory Approach: This approach was developed by the World Meteorological Organization (WMO) and the United Nations Environmental Programme (UNEP) to facilitate broad application by parties to the Framework Convention on Climate Change (FCCC). It utilizes a series of worksheets, each related to a different source of carbon flux; the worksheets contain formulas necessary to compute carbon storage. This method uses data on land use and land management changes over time, stratified by climate and soil type. Soils are merged into six classes based upon texture, morphology, and ability to store organic matter; climate is divided into eight categories based upon average annual temperature, average annual precipitation, and length of the dry season. Initial estimates have been made for U.S. officials to use in international climate change discussions (Eve et al., 2001). Acreages of land use, management, and soils obtained from the NRI were aggregated at the MLRA (Major Land Resource Area) level; there are about 180 MLRAs in the U.S., representing a geographic unit with relatively similar soils, climate, water resources, and cropping systems. The IPCC worksheets were developed using NRI data, additional information on use of conservation tillage management, and climate data that were 1961 to 1990 averages aggregated to the MLRA level. Results using preliminary 1997 NRI data showed U.S. farmland now provides a net carbon balance of 20 million metric tons per year. Initial projections are that with improved management, U.S. farms and ranches have the potential to store an additional 180 metric tons per year, which is equivalent to 12-14% of total U.S. emissions. Earlier analyses using 1992 NRI data had indicated that all regions except the Sub-Tropical Moist Region (extreme southeast U.S.) were providing a net increase in soil carbon from 1982 to 1992; the most positive results were in the Warm Temperate, Moist Region (the largest region, with soils and climate that produces high yielding crops, and where there was the highest level of adoption of no till).

(2) Erosion/Productivity Impact Calculator (EPIC) Model: EPIC is a daily time-step model that simulates weather, hydrology, soil temperature, plant growth, erosion, runoff, and leaching of water and nutrients; it includes a component for soil carbon but currently accumulations are not handled properly. EPIC will be used with the NRI database and some ancillary data to run 50-year simulations to estimate baseline conditions and to look at the potential for carbon accumulation in soils under alternative practices (Goss et al., 2001). Eventually, the EPIC model will be run for each NRI point that is in an agricultural use; but currently the information requirements are too great, and a reduced number of soil and climate groups are being used. This procedure requires nutrient management and conservation tillage management data, which are not collected directly by current NRI methodology. However, a pilot study was conducted in 1992 to determine the feasibility of collecting conservation tillage data as part of the NRI, and data were collected at 10% of the NRI cropland sites. Those data plus annual Conservation Tillage Information Center (CTIC) information are being used to derive a prediction equation for imputing tillage condition at each NRI sample point. Nutrient management information comes from the Cropland Practices Survey, conducted by the USDA National Agricultural Statistical Service.

(3) CQESTR Model: This is a new field-level model being developed by the USDA Agricultural Research Service, to allow farmers to determine short-term carbon gain or loss each year based upon specific changes in management. CQESTR describes the effects of crop rotation and tillage on decomposition of crop residues, incorporation of amendments and residues into soil organic matter, and storage of carbon in the soil. The model is sensitive to climate and soil, agronomic practices (cropping history and rotations, cover crops and tillage, fertilization and amendments, erosion), and initial soil carbon; it needs data on seasonal temperature, crop or amendment added, tillage practices (timing, incorporation), nitrogen content of residues, and soil characteristics (organic matter, bulk density, depth by layer). CQESTR is being developed to help with the formulation of national carbon sequestration policies and programs, as well as for application at the farm level.

(4) Century Model: This is a more general, long-term plant-soil-nutrient model developed by Colorado State University, in conjunction with the Agricultural Research Service, that links the carbon, phosphorus, and nitrogen cycles. It provides a comprehensive simulation of carbon dynamics across an entire ecosystem for a number of years, by considering plant responses to soil nitrogen and management practices such as no-till to predict crop yields and levels of carbon. Century results will be compared with the IPCC approach (Eve et al., 2001).

6. Summary

The NRI database provides a useful analytic framework when analyzing agri-environmental resource issues. It provides properly correlated site-specific data on soils, land use and land use change, and other factors related to conservation issues for a large number of sample sites spread throughout the country. The NRI data are linked directly to the soil interpretations database and various climate databases; they can be linked in less direct ways (geospatial methods) to many other types of data (for example, data on socioeconomic issues, or data on chemical applications).

Soil erosion is one of the major issues addressed by the NRI; the database provides considerable data that enables many forms of complex analysis of problems associated with erosion. Data from the 1982 NRI provided the basis for development of Conservation Compliance and the Conservation Reserve Program that were part of the Food Security Act of 1985 (Goebel and George, 1998). Various concepts and regulations regarding highly erodible land were developed, particularly the concept of inherently erodible land. An erodibility index (EI) was derived which took into account the physical factors used in the Universal Soil Loss Equation (USLE) and Wind Erosion Equation (WEQ). This index was used to rank inherent erodibility of any cropland parcel by assigning each a relative, normalized number. The EI concept does not take into account those erosion factors that represent human management of the land; those factors can be used to make comparisons between eroding conditions and inherent erodibility. The NRI database was used to develop possible rules and regulations for the Food Security Act and to then examine socioeconomic and political considerations. Environmental benefits dealing with water quality and wildlife became more important in the 1990 and later Farm Bills; the NRI assisted in development of various additional provisions (USDA, 1999).

Agri-environmental issues discussed in earlier sections can be addressed using NRI data just as can issues dealing with soil erosion. Models can be run for each sample point and results can be analyzed geostatistically taking into account political, socioeconomic, and environmental considerations. For issues such as soil erosion, the NRI modeling is fairly straightforward and the NRI database (including associated soils information) contains most of the needed factors. For issues like water quality and soil carbon storage, various factors must be imputed from other sources of data; such imputation must make assumptions about how to associate the external data with the NRI data. The assumptions associated with the imputation must be checked for "sensitivity"; procedures will evolve as the methodology becomes more mature and analysts become more familiar with the data sets and particulars of the models.

The use of the NRI database and its 800,000 sample points provides a scientifically-credible framework for statistically aggregating the results – by providing estimates for various portions of the universe, using the sampling weight (expansion factor) associated with each sample observation. Using each sample observation provides the "proper distribution" of all the various combinations of land use and management history, soil, and many other factors. However, a number of analyses are performed with NRI data where the data are aggregated before the models are run. This aggregation of the data before the models are run may well mask certain critical results. The most unusual situations are often the most harmful relative to environmental factors; these are in the "tails" of the statistical distributions of the NRI variates and will be lost or averaged out if only aggregate or "representative" values are used. This is an important consideration when analyzing agri-environmental issues with any type of modeling.

References

Breidt, F.J., McVey, A.M. & Fuller, W.A. (1996) Two-phase estimation by imputation, *Indian Society of Agricultural Statistics*, 49, 79-90.

Campbell, C.A., Gregorich, E.G., Zentner, R.P., Roloff, R., Janzen, H.H., Paustian, K., Smith, W., Liang, B.-C. & McConkey, M.G. (2001) Carbon sequestration in the Canadian prairies: quantification of short-term dynamics, in*: Soil Carbon Sequestration and the Greenhouse Effect*, R. Lal (Ed.), SSSA Special Publication No. 57, Soil Science of America, Inc. Madison, Wisconsin, 93-114.

Eve, M.D., Paustian, K., Follett, R.F. & Elliott, E.T. (2001) An inventory of carbon emissions and sequestration of U.S. cropland soils, in*: Soil Carbon Sequestration and the Greenhouse Effect*, R. Lal (Ed.), SSSA Special Publication No. 57, Soil Science of America, Inc. Madison, Wisconsin, 51-66.

Flather, C.H., Brady, S.J. & Inkley, D.B. (1992) Regional habitat appraisals of wildlife communities: a landscape-level evaluation of a resource planning model using avian distribution data, *Landscape Ecology*, 7, 137-147.

Fuller, W.A. (1999) Estimation for the United States National Resources Inventory, *1999 Proceedings of the Survey Methods Section*, Statistical Society of Canada (in press).

Goebel, J.J. (1998) The National Resources Inventory and its role in U.S. agriculture, in: *Agricultural Statistics 2000*, International Statistical Institute, Voorburg, The Netherlands, 181-192.

Goebel, J.J. & Baker, H.D. (1987) *The 1982 National Resources Inventory Sample Design and Estimation Procedures*, Statistical Laboratory, Iowa State University, Ames, Iowa.

Goebel, J.J. & George, T.A. (1998) Establishing a consistent national base for assessing natural resource issues, in: *Evaluating Natural Resource Use in Agriculture*, Robertson, T., English, B.C., & Alexander, R.R. (Eds.)*,* Iowa State University Press, Ames, Iowa, 335-355.

Goss, D.W., Kellogg, R.L., Berc, J.L. & Sanabria, J. (2001) A national assessment of soil carbon sequestration on cropland: description of an analytical approach, in*: Soil Carbon Sequestration and the Greenhouse Effect*, R. Lal (Ed.), SSSA Special Publication No. 57, Soil Science of America, Inc. Madison, Wisconsin, 41-50.

Kellogg, R.L. (2000) Potential priority watersheds for protection of water quality from contamination by manure nutrients, presented at *Animal Residuals Management Conference 2000*, Kansas City, Missouri.
Kellogg, R.L., Maizel, M.S. & Goss, D.W. (1992) *Agricultural Chemical Use and Ground Water Quality: Where Are the Problem Areas?* Soil Conservation Service, U.S. Department of Agriculture, Washington, D.C.
Lal, R., Kimble, J.M., Follett, R.F. and Cole, C.V. (1998) *The Potential of U.S. Cropland to Sequester Carbon and Mitigate the Greenhouse Effect*, Ann Arbor Press, Chelsea, Michigan.
Nusser, S.M. & Goebel, J. J. (1997) The National Resources Inventory: a long-term multi-resource monitoring programme, *Environmental and Ecological Statistics*, 4, 181-204.
Sorensen, A.A., Greene, R.P. & Russ, K. (1997) *Farming on the Edge*, American Farmland Trust, Northern Illinois University, DeKalb, Illinois.
U.S. Department of Agriculture (1962) *Basic Statistics of the National Inventory of Soil and Water Conservation Needs, Statistical Bulletin 317*, Soil Conservation Service, Washington, DC.
U.S. Department of Agriculture (1999) *Economic Valuation of Environmental Benefits and the Targeting of Conservation Programs: The Case of the CRP, Agricultural Economic Report Number 778*, Economic Research Service, Washington, DC.
U.S. Department of Agriculture (2000) *Summary Report: 1997 National Resources Inventory (revised Dec. 2000),* Natural Resources Conservation Service, Washington, DC, and Statistical Laboratory, Iowa State University, Ames, Iowa, 89 pages.
U.S. Department of Agriculture (2001) *Agri-Environmental Policy at the Crossroads: Guideposts on a Changing Landscape, Agricultural Economic Report Number 794*, Economic Research Service, Washington, DC.

ESSAYS

n. 12/2002

Conference on Agricultural and Environmental Statistical Applications in Rome

Rome, 4-8 June 2001

Volume IV

ISTAT

5 June 2001

Contributed Paper Session

AGRICULTURAL CENSUSES

Chair: F. Adornato

3rd Agricultural Census in Iran

Gholamali Kheiri
Statistical Center of Iran,Dr Fatemi Ave-Tehran 14144
e-mail: Alikh2002@yahoo.com

Abstract: This paper aims to clarify the actual condition of final agricultural census in Iran. This report describes the history, census objectives and uses, legal basis, statistical unit and coverage, instruction and training, holding list, data collection method and dissemination.

Keywords: Agriculture, Census, Iran, Objective, Frame, Survey, Coverage, Minimum size, Area, Crop, Sheep, Goat, Instruction, Training, Legal, FAO, List, Data, Statistic, Register, Sample, Dissemination.

1. History

The Statistical Center of Iran has conducted three Censuses of Agriculture so far.
The General Census of Agriculture 1993 is the 3rd and the latest census carried out in Iran. The first was conducted in 1973 and second in 1988. Data gathered by the censuses provide precise information on the structural aspect of agricultural holdings, including land use trends, crops production, livestock inventory and other information on such activities as apiculture, sericulture, pisciculture and machinery use.

2. Census objectives and uses

Two basic objectives of the census were:

a- To provide a frame for other agricultural sample surveys,

b- To provide data for small administrative units.

The census of agriculture is the main source of statistics about the nation's agricultural production and the only source of data at each geographic level. Census Statistics are used by the Government and Islamic Consultative Assembly to develop and change the form of programs and plan for the future. Many national and provincial programs use census data to design and allocate funds for the extension of service projects, agricultural research, as well as in universities. Private sector uses census statistics to provide more effective production.

3. Legal basis

The 3rd General Census of Agriculture was carried out according to an order endorsed by the President. This census was taken, as far as possible, according to the recommendations for decennial programs promoted by FAO.

4. Statistical unit and coverage

The agricultural holding is the statistical unit for the census. An agricultural holding is an economic unit of agricultural production under single management without regard to title or legal form. Grazing land, forest land and all land not tilled by agricultural holdings are excluded in the census.
The census covers agricultural holdings nationwide , including both rural and urban areas. In Iran, a minimum size limit is adopted for holdings, only into the following major groups:

- $400m^2$, Area for temporary crops;
- $200m^2$, Area for permanent crops;
- 2 heads of sheep or goat;
- 5 chickens or ducks .

5. Instruction and training

Personnel assigned to census operations were provided with a booklet entitled "Instructions for census enumerators", which describe the enumerator's duties and includes definitions regarding survey units, the field of observation and each item of on the questionnaire.

6. Holding list

The list of all holdings was provided during enumerating operations. There were not any model from administrative or registered sources. Agricultural holder Households were identified by visiting their ordinary place of residence and their holdings were enumerated. Holdings operated by legal persons were enumerated by visiting directly the related places.

7. Data collection method

This census was implemented by enumerators who visited sources of information in a direct interview method. Household members, agricultural holders and local well-informed persons were interviewed respectively in order to complete Listing Form, Holding Questionnaire and Village Questionnaire.

8. Data dissemination

The system for the dissemination of data from the 3rd General Census of Agriculture 1993, provides the appropriate publications. The types and levels of data analysis that will be made available to users are described in the book "Detailed Result of General Census of Agriculture 1993" and more specific data are available on floppy disk according to standards which will allow users to undertake further research.

References

Statistical Center of Iran (1993) Collection of Forms and Instructions, *Documents Of Central Committee of Agricultural Census of Iran 1993*.

Yugoslav Census of Agriculture 2001 as an Instrument of Supplement in Agricultural Statistics

Vidosava Lukic
Federal Statistical Office, Kneza Milosa 20, 11 000 Belgrade, Yugoslavia
e-mail: Vidosava M Lukic@szs.sv.gov.yu

Abstract:The aim of this paper is to present the role of Yugoslav Census 2001: classification by different characteristics, obtaining data for small territorial units and testing data accuracy.

Keywords: Census, Agriculture, Statistics

1. Introduction

Censuses of agriculture in Yugoslavia were conducted in 1931, 1950, 1960 and 1969 Due to financial and rational reasons there were no separate census in 1981 and 1991, but data on agriculture were collected along with censuses of population, households and dwellings. Also the associated census has been carried out in 2001 as "Census of Population, Households and Dwellings 2001". One of the Census units is an agricultural holding, the other ones are person, household and dwelling. There are 16 questions in the Census referring to agriculture. These questions associated with other questions on persons and dwellings obtain information on agrcultural holdings and their main characteristics.

2. Census 2001 Supplement

2.1 As in Yugoslav statistics there are no farm structure surveys at all, decennial agricultural census is an instrument which obtains specific data on agricultural structures. Priority of analytical needs is fulfilled on the base of Census data classified by holdings grouped according to different characteristics. The aim of the Census is not to obtain aggregate total data on capacity and production, because these data exist and are collected through current statistics. Aggregate totals on land area, livestock etc. are in the secondary priority of Census. Nevertheless these data serve to provide data for small administrative units and as benchmarks for testing accuracy of some data of current statistics.

2.2 Analyses of the structure of agriculture and its most important phenomena are performed in the Census through a series of classifications: by size of total and areable area of holding, tenure of land, source of income, labour input. 1) Classification acording area magnitude of land is traditinal one. It is done according to a) total area used, b) areable land used and c) total area owned. Holdings with different structures of land and production are classified in the same groups in this way. Also the lack of such kind of groups is caused by the fact that livestock does not determine size of holdings. The complex census wich pretends to cover three censuses (population, dwellings, agriculture), as it did in 2001, with priorities to obtain valid information on population, makes it impossible to extend it with more questions from agricultural domain which would permit to form better criteria for measuring the size of holdings. 2) The studies from previous censuses confirmed that classification by tenure is a very useful tool in the analysis of agricultural holdings. These classes are a) holdings which neither lease land to others nor rent from others; b) holdings which rent the land from others; c) holdings which lease land to others, with a further breakdown of the last two classes according to the percentage of area rented. 3) The aim of classification according to age of holding members – agricultural workers (holdings with members up to 60 years and with members above 60 years) is to obtain data about "old holdings" and thieir characteristics. The classification is useful for the analysis of the process of "getting older" and "die away" of agricultural holdings. 4) Classification according to socio–economic types (agricultural holdings, nonagricultural holdings, mixed holdings, holdings without workers) can give a sufficiently adequate picture of rural disintegration or integration. The class of so–called "mixed" holdings has shown that it comprises qualitatively different holdings. 5) Classifications according to number of holding members are done according a) total number b) number of economically active c) number of agricultural workers d) number of permanent agricultural workers and e) number of women–agricultural workers. These classifications can provide an adequate analysis of agricultural "pressure" and labour force supply on holdings. For analytical purposes all these classification are combined with census items and with each other.

2.3 Censuses in Yugoslavia, particulary the present one in 2001, are the only source which insure data on livestock number for small administrative units. The data on the number of livestock are obtained throught current statistics by regular annual reports. Reports obtain data for bigger administrative units, while only the Census offers data for the smallest territorial units as settlements and statistical enumeration areas. Data on livestock for small territorial units, obtained in a decennial period from censuses, serve, beside others, as a frame for planning and setting samples for livestock surveys.

2.4 Last years Yugoslav statistical data on production obtained from regular reporting servicies, especially for wheat, have been critizied from the cabinet ministerial level. Financial reasons made impossible for agricultural statisticians to undertake serious steps to resolve this problem. For years there have not been objective measurements of areas and production in Yugoslavia. Having in mind that data on wheat and maize are very important (in total value of agricultural production they comprise about 30%), in the Census 2001 questions are included on harvested area, production and stocks of wheat and maize. The intention is to test current statistical accuracy by comparing data on wheat and maize from the current statistics with those from Census. As far as it is known, data on capacity from censuses are lower than what is actually the case. Therefore data on wheat and maize obtained from Census 2001 may serve as a correction on current estimates for those territorial units in which census data are significantly higher.

3. Conclusion

The current statistics is aims at giving the annual data on capacity and production in the form of totals for Yugoslavia as a whole and by regions. On the other hand the census 2001 is needed for decennial analyses of holdings structures through series of classifications and relations of basic economic and social characteristics and also for providing data for small territorial units and testing current statistics accuracy.

The Norwegian Census of Agriculture 1999

Ole Osvald Moss
Head of division, Primary Industry Statistics, Statistics Norway,
N-2225 Kongsvinger, Norway, e-mail: ole.osvald.moss@ssb.no

Dagfinn Sve
Project supervisor, Census of Agriculture 1999, Statistics Norway,
N-2225 Kongsvinger, Norway, e-mail: dagfinn.sve@ssb.no

Abstract: Contrary to earlier censuses, the 1999-census was conducted without interviewers, and the data were collected both by postal questionnaires and by screening of administrative registers. Three months after receiving administrative data, preliminary figures on livestock and use of agricultural land were published at municipality level on the Internet. Early results from the main questionnaire were calculated by drawing a sample from the total population. Final results have been published free of charge at low regional levels on the Internet.

Keywords: Census, agriculture, Norway, publishing, data sources, registers, administrative data, results, Internet.

Questionnaires and administrative registers

Three different questionnaires and several registers were used to collect data. Every respondent had to fill in the main questionnaire. Respondents who did not apply for governmental grants at the National agricultural authority (NAA), had to fill in an additional questionnaire about area use and livestock. Finally about 5.000 respondents had to fill in the Greenhouse nursery and horticultural form. In July 1999 census questionnaires were mailed to about 83.000 farmers who were registered with any type of activity (production, application for grants etc.) in the governmental Farm Register (FR).

Table 1: *Data collected from which data source*

Themes	*Data sources*
Use of agricultural area	NAA Register, SN Additional questionnaire.
Livestock	NAA Register, SN Additional questionnaire
General information, (ownership etc.) land areas (ditching, reclaiming, use of artificial fertilisers etc.)	SN main questionnaire
Farm buildings	SN main questionnaire
Machinery and implements (Contracting work and machinery pooling in last 12 months), irrigation systems	SN main questionnaire
Education	SN main questionnaire and Register
Labour input, supplementary industries	SN main questionnaire

Income and property	SN Register
Vegetables, fruits and berries. Greenhouse production and green-house equipment	SN Greenhouse nursery and horticultural questionnaire

The farmers report data to the NAA on agricultural land use and livestock twice a year to apply for governmental subsidies. Statistics Norway (SN) has worked closely with the NAA to adjust the application form in order also to cover statistical needs. The unique identification key in the FR is a farm number consisting of municipality-number and a property-number. In addition to this, each farmer's birth number was used to link the data sources.

Data on education, income and property of holders and spouses are obtained from other registers within SN. All together, this combination of production data, labour input data and economic data caters for unique possibilities for in-depth analyses.

Preliminary figures and final results

Administrative data on area use and livestock supplemented by preliminary data from the "additional questionnaire" made it possible to present preliminary statistics at low regional level already in February 2000 on the web: http://www.ssb.no/jt1999_en/.

A sample of representative municipalities were picked out and edited first. This sample, containing 15.000 respondents, was stratified by size of agricultural area in use, number of dairy cows and size of area of grain and oil-seeds. Estimates were made to present statistics on regions of two or more counties. An "early harvest" of articles with preliminary figures on various themes, was displayed on the Internet in March and April 2000, and received a lot of attention from various users, in particular from media focusing on agricultural and regional affairs.

The publishing of final figures on municipality level on the Internet, was performed county by county from December 2000 to February 2001. Most of the results are published on the Internet without any costs for the users. Text, graphs and thematic maps accompany detailed tables. Simplified maps are also used to guide the users by clicking on the actual county or municipality.

In addition to the Internet, 19 publications on paper are prepared spring 2001 - one publication for each of the18 counties with figures on municipality level, and one "over-all" publication with figures for counties and the whole country. This last publication will be published in an English edition. A separate analysis on the economy of agricultural holdings will be presented summer 2001.

The main users will receive more complete information about the census material and the possibilities of linking to other data sources. It is a goal that the census staff will contribute with articles oriented towards media - both "common" newspapers and professional agricultural magazines/bulletins etc. Eventually, this can lead to more attention to the census material, and some users are expected to pay for tailor-made statistics.

Using Classification Tree Methodology for Preliminary Quality Evaluation of the 2000 Census of the Agriculture List

Antonia Boggia
Italian National Institute of Statistics; Via Depretis, 74/b, Rome, Italy,
e-mail: boggia@istat.it.

Alessandro Pallara
Italian National Institute of Statistics; Via Depretis, 74/b, Rome, Italy;
e-mail: pallara@istat.it.

Keywords: list frame development, Classification Trees, coverage, administrative data.

1. Introduction

The census of agriculture, taken every ten years, collects data and publishes information on land usage, crops and livestock, and operator as well as farm characteristics of farms in Italy. A *census farm* is defined as a place where any crops were grown or any livestock or poultry were raised or any woodland was operated during the Census year. Any farm size or kind of farm operator is eligible for inclusion in the farm census list, except for agricultural operations taken on very small parcel of land for home-owner consumption.

Developing the list of farm operators for the 2000 Census of Agriculture has been a particularly complex process that involved merging and unduplicating several source lists of names and addresses believed to represent agriculture operations, including the previous agriculture census list, and administrative records of the Ministry of Finance, the Ministry of Agriculture, other government agencies and trade associations. The list compilation was carried out in two phases. During the first phase, records from the various lists were linked on names, addresses and tax return code with the purpose of eliminating records relating to the same farm. This linkage operation yielded a preliminary list of about 4.7 million names and addresses, which were separated in two files, one including all units enumerated in the previous census of agriculture (about 3 million farms) — which had either matched or not with any of the records in the other source lists — and one including records from auxiliary sources non-matching with any record in the previous census list but matching with one another in different lists. Indeed, because of heterogeneity of the units in different source lists (individuals, businesses with multiple establishments, large organizations having some association with agriculture) and problems with the matching variables, it was suspected that the preliminary list still included many potential duplicate records. Therefore, a second phase of the list development was carried out, consisting primarily of verifying the information for each record in the preliminary list, to ascertain which one corresponded to actual farm operations. This represented a very complex and costly task to be accomplished by local census offices who have the responsibility of field work during census operations. After completing this process a final list was created, containing approximately 2.75 million names and addresses used for census direct data collection.

2. The classification tree model and the list quality assessment

A field enumeration survey has been planned at the end of census data collection, involving an independent measure of the completeness of farm counts based on an area frame. However, because of several innovations introduced in the procedure for the 2000 census list development (mainly due to extensive use of administrative records) it was desired to obtain some preliminary indications about the quality and accuracy of census list. Therefore, in April 2000, a *farm identification pilot survey* was conducted on a sample selected from the preliminary list (i.e. the one including 4.7 million names) obtained after the first phase of census list development described above. The purpose of this survey was to identify list records that did not represent census farms and potential duplicate records, mostly due to the heterogeneity of units in the source list. A statistical model has then been developed, which classifies records into groups of probable farm and non-farm operations using classification and regression tree methodology (CART, Breiman *et al*., 1984). The classification tree model was constructed combining survey results and auxiliary information in the source list (farm size, kind of ownership, number and type of administrative record sources). The tree model defines an optimal classification criterion for recursive partitioning of the units in the list into subgroups which are maximally internally homogeneous. The classification tree methodology has been applied in recent agriculture censuses in United States to reduce the number of names and addresses in the mail list (Owens *et al*., 1989; Schmehl and Ramos, 1990; Bureau of Census, 1996).

The farm identification survey data served as the source for the classification tree definition. Based on the tree selected, groups of records with similar characteristics were created and assigned a probability that a unit in the group was a farm. This probability was estimated in each group through the proportion of observed farms at the survey. The tree model and the estimated probability could play an important role in different operations related to the census of agriculture:

- the tree model may be used for profiling each record in the preliminary list (namely, defining the group to which it belongs). Each record will then be predicted as having a farm/non farm status. This predicted value will be compared with the actual status identified at the end of the second phase of the list development, shortly described above. A comparison, for each model group, between the predicted farm proportion and the actual farm counts resulting from the second phase of list development could indicate accuracy problems with the latter operations and a way to improve coverage of census list. Although this approach has found only limited applications for 2000 Census of Agriculture, it could be implemented in future censuses, in order to reduce the cost of list development (note that one sixth of the total budget of the 2000 Census has been spent for the second phase of list development). Namely, for completing the census list, after the preliminary matching operations among the various source files, statistical modeling could be employed to identify and delete from the file records believed least likely to represent farms;

- the tree structure defined could be helpful to discriminate - among the various administrative sources used for constructing the preliminary list - those which are more useful to classify records between farm and non-farms. This will give valuable information for updating an agriculture list frame during inter-censual years using mainly administrative sources.

References

Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984), *Classification and Regression Trees*, Wadsworth International, Belmont, CA.

Bureau of Census (1996), 1992 Census of Agriculture. Volume 2: Subject Series, part 4, History, *report AC92-S-4, U.S. Department of Commerce, Washington*.

Owens D., Killion R.A., Ramos M., Schmehl R. (1989), Classification Tree Methodology for Mail List Development, *American Statistical Association, Proceedings of the section on Survey Research Methods*, 593-598.

Schmehl R.L., Ramos M. (1990), Evaluation of the Classification Tree Methodology Used for the Development of the 1987 Census of Agriculture Mail List, *American Statistical Association, Proceedings of the section on Survey Research Methods*, 308-312.

5 June 2001

Contributed Paper Session

AGRICULTURAL STATISTICAL SYSTEMS

Chair: Y. Ivanov

Spanish Agricultural Statistics

Mariano Ruiz-Espejo
UNED and UPSAM, Apartado 19207, E-28080 Madrid, Spain
e-mail: ruizespejo@terra.es

Abstract: Some data about the actual Spanish agricultural censuses and statistics are offered here regarding qualifications of the workers involved, organization and objectives.

Keywords: Agricultural censuses, Agricultural statistics, Explotations, Objectives, Organization, Sociology, Spain.

1. National Agricultural Statistics

The fifth Spanish agricultural census was realised in 1999 by the INE (Instituto Nacional de Estadística, National Statistical Institute of Spain), after previous censuses in 1962, 1972, 1982 and 1989.
The census, and other intermediate specific surveys between two consecutives censuses, constitute (with other statistical operations run by the Spanish Ministry of Agriculture) the agricultural information with an official character produced in Spain; further Eurostat publishes agricultural economic accounts and prices of the 15 countries which are actual members of the European Union.
To facilitate the agricultural evolution and series' comparisons, the last census covers higher number of cultivations than those proposed in the Eurostat list.
All the Spanish agricultural activities, on September 30 (1999), are investigated and observed over the period of October 1 (1998) to September 30 (1999), which is the data's reference period.
About 2.5 million of exploitation holders have been interviewed in approximately 7000 municipal census offices in Spain. The census also investigates the methods of ecological and environmental agriculture in the exploitation.
The sample size, for additional observations and derived from the previous directory of 2.5 millions of holders, was 1080 units or holders. The census directory was interviewed by 6000 census agents; they also had 1500 group representatives and 230 inspection assistants for information quality control.

2. Specific Spanish Autonomous Communities

Cataluña and Euskadi, two important "Autonomous Communities" of Spain, have employed other staff for their contributions to the Spanish Agricultural Census 1999, produced by the IDESCAT (Institut d'Estadística de Catalunya) and the EUSTAT (Euskal Estatistika-Erakundea).

Concretely, in Cataluña, they had 265 agents and 90 additional coordinators. In Euskadi, they had 5 technicians, 19 inspectors and 150 interviewers. Cataluña had 34 agricultural district offices, and Euskadi had 21 of them.

3. Present and Future

The Spanish agricultural census directory is the frame for complementary inter-census agricultural surveys, and for the future agricultural census of the year 2009.
For the 1999 census, there have been 52 provincial delegations of the INE, and 241 "territorial districts" (comarcas territoriales) in the country during the period mentioned.
Since 1989, the Spanish agricultural censuses have been adapted to the census methodology of the European Union, and it represents the main contribution of the Spanish agricultural statistics for internal use and for the normalized use in the European Communities via Eurostat standards or agreements.

Acknowledgements

Some data of this article have been obtained from the reports by A. Morate, L. Alonso, C. Ballano, J.M. Bercebal, M.D. Lorca, P. Sánchez, J. Capellades and M. González, in the review *Fuentes Estadísticas* No. 40, December 1999 / January 2000, pp. 13-19, which was published in Madrid.

Methodological Problems of Food Balance Sheets Construction in Russia Using FAO and Other International Organizations Experience

Anatoly Manellya, Marina Lezina
Centre for Economic Analysis under the Government of the Russian Federation, Russia
e-mail: postman@cek.rfnet.ru

The food balance sheets system existing in Russia at the present time represents one of the most important statistical instruments and it is widely used for analysis and forecasting of law-governed characteristics of agroindustrial production development. The estimation of the system from the standpoint of modern economic conditions, testifies to the necessity of its improvement. The end of the centralized distribution of agricultural production, the liberalization of its market, the non-state retail trade development of food-stuffs, the stimulation of interregional relations within the agroindustrial complex through commercial intermediary services, presents a series of tasks to statistics which cannot be solved without introducing specific changes in the actual food balance sheet methodology in Russia.

A special urgency of the problem is connected with the fact that the opening of the Russian economy to the world market demands an adaptation of food balance sheets to international standards using the experience of FAO, OECD, Eurostat and other international organizations. Without that it is not possible to make a correct comparison between the most important economic and statistic indices of Russia and the data of foreign countries. At the present time, a limited level of adaptation is not enough and in connection with this fact, an approach to achieve the above-mentioned balance construction has to be transformed as far as possible in accordance with international statistics principles.

The improvement of food balance sheets construction methodology intends to solve a series of methodological problems indicating below.

- To make a food balance sheets structure more precise. To substantiate an inclusion in the balance composition of items: "Industrial processing of agricultural production for technical purposes" and "Industrial processing of agricultural production for feeding purposes".
- To provide calculations of the item "Fund of personal consumption" by two methods: by a basic principle and by a final product. To elaborate methodology and an algorithm of the balance item calculation to a final product and also a scheme of correction results obtaining by the two methods.
- To make a food balance sheets nomenclature more precise. To arrange stages and a succession of the construction of this balance.
- To make the definition of terminology in food balance sheets construction more precise. To compile a dictionary of corresponding terms.
- To make the estimate of food balance sheets structural components more accurate. To draw up an inventory of applied calculation methods and to determine the weakest points in their methodology and they results. To discover and enumerate indices when determining various valuations (including experts'). It is necessary to

allocate particularly the balance elements estimated with the aid of a final counting method. To determine methodology with the purpose of increasing the basis for estimated calculation.

- To elaborate a methodological approach and a scheme of establishing the correlation of food balance sheets on the federal and regional levels.
- To concretize budgetary surveys by programme changes. This is a necessity that appears in connection with an improvement of the food balance sheet construction methodology.
- To draw up an inventory of coefficients applied to basic product for food balance sheets construction. To replace obsolete coefficients by new ones, and to establish their quantity levels.
- To clear up peculiarities of regional food balance sheets construction, and their differences from balances of the federal level. To pursue a reasoning of regional food balance sheets construction as a separate original task.
- To prepare suggestions for an organizational structure with the purpose of co-ordinating and systematizing activities in food balance sheets construction. To compose a list of organizations taking part in that kind of work, to make information and methodological communications between them more precise and to provide suggestions on their improvement.
- To work out methodology and algorithms for the calculation of value food balance sheets.
- To work out methodology and algorithms for the calculation of food balance sheets for the future.

Methodological Approaches to and Analytical Applications for the Advanced Surveys of Agricultural Producers in Russia

Mikhail Kozlov
Centre for Economic Analysis under the Government of the Russian Federation, Russia
e-mail: postman@cek.rfnet.ru

The main purpose of sample surveys carried out by the Centre for Economic Analysis consists of obtaining new additional information in order to characterize the economic and financial situation of agricultural producers considering that such information is absent in the all-round statistical registration of Russian Goskomstat. In addition, a higher effectiveness of obtaining information is achieved with the help of sample surveys to the extent that detailed all-round statistic registration data are generalized at the country's level, for example for the preceding year in 8 or 9 months only.

From 1992 to the present advanced surveys of the economic and financial situation of agricultural producers has been carried out by the Centre twice a year. Most of the surveys were conducted together with Goskomstat of Russia and, since 1998 – jointly with the Federal Agency of Governmental Communication and Information under the President of the Russian Federation (FAGCI). Next to this, functions of our Centre have consisted of an elaboration and co-operation of a survey instrument set (questionnaires, instructional materials to fill in forms), preparation of the form and content of final documents and analytical materials (notes, reports, etc.) for the Government, ministries, departments, scientific centres and institutes.

Analytical materials received as a result of sampling survey information processing reveal the following characteristics: a development of agrarian sector reforms, agricultural producers intentions to enlarge or decrease the basic types of agricultural production, economic situation of the producers, their adaption to market conditions, directions out of financial crisis, etc. The materials are used in the long run to establish governmental policy in agrarian sector development.

An area and farm representative sample was made together with Goskomstat and FAGCI. The agricultural enterprises and individual (peasants') farms of regions representing basic natural, climatic and agricultural zones of Russia were included in the sample survey.

A task of dissemination of the obtained analytical quantity results of surveys on the country as a whole has not been carried out in recent years, taking into account the large diversity of soil and climatic conditions and instability of the production results during the transition period in Russia. Consequently, quality indicators have been mainly included in questionnaires of recent surveys. The approach has made it possible to receive a fairly reliable estimation of the main tendencies in the economic position of agricultural producers and of the agrarian sector in Russia as a whole during the transition period.

Some conclusions based on the recent survey of economic and financial position of agricultural producers in 1999 are given below. The survey was carried out in 38 regions of Russia in January 2000 (1522 questionnaires were obtained and processed).

An analysis of response data provided the information which so far was absent in all-round statistical registrations. Thus, about 30% of all respondents desired an increase in regional interference in agrofood markets with the help of restrictions and an embargo of transport of agricultural production beyond the regions border. Such measures lead to an extension of shadow production and other economic violations. In the opinion of 16% of all respondents, including 24% of the farmers, the volume of shadow cereal output in their regions (farms) exceded 20%.

About two thirds of our respondents drew attention to a drop in the production of cereal stocks by the rural population, 53% - of meat and meat products, 44% - of fruits, etc. At the same time a considerable part of respondents (42% and 28%) considered that the production of vegetables and potato stocks by the population remained in 1999 at the 1998 level. An important part in food provision for the population comes from private subsidiary holdings and garden plots of the population at the expense of which (in the opinion of 44% of respondents) more than 50% of all food-stuffs consumption are provided in their regions (farms).

Owing to a lack of financial resources and a reduction in material and technological provisions for the agrarian sector, priority state support has to be rendered to the strongest producers which could survive in the difficult conditions of the transition period in the following forms: allocation of short duration cheap credit (in the opinion of about 53% of respondents); a restructuring of agricultural producers' debts to the state (30%); state participation in expenditures by introduction of new technologies in agriculture (25%); organization of co-operatives on agriculture production purchasing private subsidiary holdings of the population and farms (nearly 25% of the respondents).

It has to be underlined in conclusion that the materials received indicate an increasing adaptation by agricultural producers to the market conditions in Russia.

Recent Developments in the Crop Forecasting System of the MARS Project of the JRC-EC(1)

Giampiero Genovese
Joint Research Centre, EC, Space Application Institute, Ispra (VA) Italy,
e-mail: giampiero.genovese@jrc.it

Abstract: One of the activities of the MARS project (Monitoring Agriculture with Remote Sensing) of the European Commission, is to provide short term forecasts on crop harvests at pan-European level. Specifically the MARS crop yield forecasts are based on agro-meteorological data as derived from satellite observations and meteorological ground stations and integrated by yield time series. Recently the MARS crop monitoring and forecasting system has been improved by integration of CORINE/NDVI data (Noaa-Avhrr/Vegetation) and by the experimental use of data from weather numerical models. The MARS crop forecasting system has been extended on 4 pilot area: Russia and CIS Countries, Mediterranean Bacin, Horn of Africa and South America.

Keywords: Crop Yield forecasts, NDVI, global circulation models

1. Brief introduction of the MARS Crop Yield Forecasting System (2)

The MARS project runs since 1993 a Pan-European short term forecast activity about crop growth and yields condition (Vossen & Rijks, 1995). The activity has been carried out within the Agriculture and Regional Information Systems (ARIS) Unit of the Space Applications Institute, Joint Research Centre of the European Commission. It is based on three different kinds of information: indicators derived from meteorological data, agro-meteorological simulated indicators at crop level and indicators from earth observation instruments. The indicators produced are used as crop yield predictors where the forecasting model is fitted using the official national crop yield time series. The modelling also introduces a trend factor. Agro-meteorological scenarios are produced and analysed to describe the possible evolution of the season from the moment the forecast is issued to the crop harvest. The analyses and the main results are published into the MARS bulletin (http://mars.aris.sai.jrc.it/stats/bulletin).

2. Data, Methodological and Extension Advances

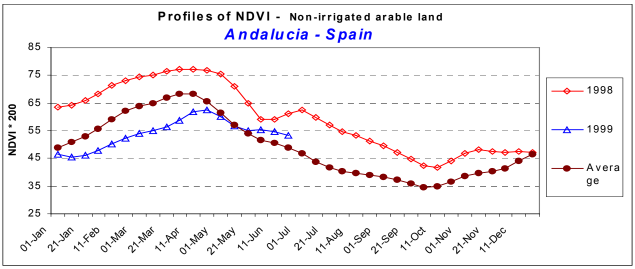
The main advances are linked to the use of new data within the system which reinforce/improve the initial measures in terms of reliability and availability of the

(1) The present paper is financially supported by The European Commission

information. These are: the introduction of data from the numerical global circulation models (ECMWF) to substitute the interpolated meteorological observation at station level (Genovese & Terres, 1999), and the introduction of the Spot 4/Vegetation data. The meteorological model data are currently used in re-analysed form and for areas of interest outside Europe. The advantages rely in the possibility of augmenting the skill of the meteorological layer where data are missing or of poor quality. Experiences are also collected for the introduction in the system of the seasonal forecasts and to solve recluted issues like downscaling of the information at more local level. In the field of low resolution satellite information new data are coming from the sensor Vegetation on board of Spot 4. The advantage consists in the possibility of using data with a higher atmospheric and geometric accuracy as far as the use of NDVI (Normalised Difference Vegetation Index) is concerned.

The main methodological contribution is given from the integration from two Pan-European data sets: CORINE land cover and the NDVI data (Genovese *et al*., 2001). The integration allows to obtain more refined class-cover profiles at the regional level, to better exploit the NDVI information for agricultural assessment and crop estimation.

Figure 1: *NDVI Profile extracted on the non-irrigated arable land class (Courtesy MARS bulletin, Vol.7 n°3).*



The methodology developed is being extended to new areas of interest of the MARS System: Russia and CIS countries, the whole Mediterranean Basin, The Horn of Africa (IGAD countries) and South America (Mercosur area).

References

Genovese G., Vignolles C., Negre T., Passera G. (2001): Combined use of NDVI and CORINE land cover data for yield monitoring and forecasting – A case study on Spain, in *Journal of Agronomy and Environment*, EDP Sciences, 21, 91-111.

Genovese G. & Terres J.M. (1999): Crop yield forecasts derived from agro-meteorological models., in *Proceedings of the Workshop on the application of Seasonal-timescale Weather Forecasts*, 14-15 June. ECMWF, Reading. 35-41.

Vossen, P. & Rijks, D., (1995, third print, 1996). Early crop yield assessment of the EU countries: the system implemented by the Joint Research Centre. *EUR Publication* N° 16318 of the Office for Official Publications of the EC. Luxemburg, 182 pp.

Situation in Agriculture in the Fr Yugoslavia at the Beginning of the New Millenium

Dragana Marković
Federal Statistical Office, Kneza Milosa 20, 11 000 Belgrade, Yugoslavia,
e-mail: filippi@szs.sv.gov.yu

Abstract: This paper presents situation in agriculture production, crop production and livestock in FR Yugoslavia, trough statistical data.

Keywords: agricultural production, agricultural area, wheat, livestock.

Some indicators of agricultural development:

- Agricultural production in the FR Yugoslavia, according to statistical data, has a share of 23% in the total national production;
- The share of agricultural population in total population by 1991 census is 17,3%;
- The number of agricultural holdings by 1991 census was 1 057 278;
- Index of physical volume of agricultural production:

1955=100

1980	1985	1990	1995	1999
248	248	243	234	242

- The private sector owns 85% in the total cultivated land;
- Out of total agricultural area, the cultivated area covers 78%, while 59% is arable fields and gardens;
- Structure of sown areas:

%

Cereals	67
Industrial crops	9
Vegetable crops	10
Fodder crops	14

- In the individual sector there are 10 ha of cultivable area per one tractor and in the state-owned sector 59 ha.

- Consumption of mineral fertilizers per ha of arable land amounts to 74 kg.

- Yield of major crops:

ton/ha
(per tree, vine, kg)

	1980	1985	1990	1995	1999
wheat	3,68	3,79	4,38	3,42	3,29
rye	1,53	1,57	1,95	1,80	1,69
maize	4,63	4,13	2,63	4,25	4,84
sunflower	1,73	2,34	2,11	1,74	1,48
tobacco	1,19	1,16		1,25	1,65
sugar beet	41,5	41,9	37,1	27,4	41,29
potatoes	9,6	8,97	7,10	8,49	8,03
plums	10	7	8	5	9
apples	19	13	16	12	14
grapes	1,2	0,7	0,59	0,9	0,5

- livestock and poultry:

thous. heads

	1980	1985	1990	1995	1999
cattle	2367	2438	2168	1950	1831
pigs	4561	4380	4325	4192	4372
sheep	3158	3113	3007	2671	2195
horses	222	153	98	96	76
poultry	24358	25344	29907	25591	26492

- products of livestock breeding:

	1980	1985	1990	1995	1998
meat, thous.ton	513	640	637	547	575
milk, cow's mill. lit.	1796	1974	1774	1888	2131
wool, tons	4998	4530	3990	3878	3422
eggs, mill.	1809	1903	1692	1805	1812

- pork covers 55%, beef 24%, poultry meat 15% and mutton 6% of total annual meat production;

- annual fish catch is about 7 500 tons.

Characteristics and Development of Agricultural Statistics

Slobodan Raonić
Federal Statistical Office, Kneza Milosa 20, 11 000 Belgarde, Yugoslavia
e-mail: filippi@szs.sv.gov.yu

Abstract: The system of agricultural statistics in Yugoslavia is: comprehensive, based on unique methodology, and oriented towards securing data from the level of municipality to the level of republic. It is harmonised with general statistical system and international standards. However, it was necessary in the past, as it will be in the future, to modernise the system. To that aim the results of the 2001 Census of Agriculture can contribute, as is also the case for the experimental results of the new statistical survey titled "Survey on Agricultural Holdings".

Keywords: System of agricultural statistics, statistical methods, experiment

The system of agricultural statistics, as a specific field of applied and theoretical statistics, functions within the unique statistical system of Yugoslavia. Together with the strengthening of the material position of agriculture and increased demand of society for information on agriculture, modern methodology was developed, as was also the case for the organisation of statistical investigations. Development of adequate methodological solutions, besides the specific characteristics of agriculture, was affected by other factors as well, especially by the possibility of society to put aside some funds necessary for statistical investigations.

The system is characterised by the following aspects: 1) it is complete and it covers all important phenomena and processes in agriculture, 2) it is based on a unique methodology for the country as a whole, 3) it is basically oriented towards securing data on the level of municipality, province, republic and the country as a whole, 4) from a variety of applied methods and efforts, to select the best ones to fit the conditions and subject of research, 5) very diversified subjects and ideas maximally harmonised with general statistical system and international standards.

Indicators obtained by regular and occasional surveys in agricultural statistics, and partly in other statistical fields are the following: 1) agricultural funds, 2) agricultural population and labour force in agriculture, 3) production and natural balances, 4) prices, bonuses and contributions, 5) national income, reproductive consumption, personal consumption, costs of production, 6) agricultural holdings and households.

The specificity of agriculture as an organic production method and an agrarian structure resulted in the application of various statistical methods of data collection. These methods are predominantly based on objective and direct measurements and registration of phenomena, as also on subjective methods (estimations, calculations, use of balancing methods etc.). Almost all known theoretical methods of statistical observation of a phenomenon such as censuses, surveys, reports, estimates, monographs and calculations have been and are presently used in the current system of agricultural statistics.

Data on crop production and livestock breeding from agricultural enterprises and agricultural co-operatives are collected by report method, which is based on accounting evidence. Estimation method (through a network of estimation areas and professional statistical estimators) is used to obtain data on agricultural areas and crops for rural agricultural holdings. Survey methods (with a sample covering 2% of agricultural holdings) are used to obtain data on number of livestock and increase of livestock. Measured by physical volume of total agricultural production, it is evident that data obtained by the report method constitutes a total physical volume of 18%, the share of data obtained by the method of subjective estimations is approximately 46%, while the data obtained by surveys is 36%.

As mentioned above, a large number of indicators in Yugoslav agricultural statistics is obtained by the method of subjective estimation. The level of reliability of the collected information, considering the big number of units, small holdings and the almost complete lack of data on agricultural holdings, presents a central problem which has not yet been systematically studied and solved in satisfactory way. Constant doubt about the accuracy of this information and our inability to verify scientifically and solve the problem of subjectivity in estimations, presents a central problem of statistics of agricultural areas and yield.

The system of agricultural statistics, although fundamentally developed and comprehensive, still has certain voids and imperfections. They are predominantly present in deficient coverage of certain phenomena and processes and reliability of the information obtained by the method of subjective estimates and sample method. Consequently, in the forthcoming period, our efforts should be directed towards securing extensive and relatively accurate information necessary for current economic and developmental policies in the field of agriculture.

As a result, a new statistical survey entitled "Survey on Agricultural Holdings" is due to be introduced in the system of agricultural statistics. The survey will be carried out on a sample basis, with the aim to: 1) secure data on agricultural holdings not covered by existing statistical surveys, 2) check the accuracy of information obtained through subjective estimations, 3) examine the possibility and verify the advantages of obtaining data on areas and yield of crops through sampling of agricultural holdings, contrary to data obtained through comprehensive cadastral records, 4) secure data on the use of materials, tools and labour force in the production of agricultural products.

In its initial phase, the survey will have experimental character. It will serve as control on the accuracy of information obtained by estimation, by checking the method of subjective estimation itself. If it turns out that the survey can procure information of better quality than the ones obtained by current surveys, this survey would become the basic research tool in agricultural statistics. In this manner, the number of investigations currently carried out would be reduced as well as the costs, while the volume and quality of information would be increased. However, if it turns out that data obtained by subjective estimations, which have a long-standing tradition here, have a satisfactory quality, the existing method would be carried on, with some revisions, and the sample of territorial units would be introduced in order to reduce the costs, and increase the quality of information.

The State of Agricultural Statistics in Lesotho

Mahatlane Elsie Molatoli
Bureau of Statistics, P. O. Box 455, Maseru, Lesotho, Southern Africa
e-mail:memolatoli@ilesotho.com

1. Agricultural Statistics in Lesotho

The aim of this paper is to provide information on the methods used to collect agricultural statistics in Lesotho. Agricultural Statistics started in 1949/50 when first Agricultural Census was executed by the Department of Agriculture (now Ministry of Agriculture). The second census was in 1959/60, still performed by the department of Agriculture. In 1965 when the Bureau of Statistics was established and its first agricultural census took place in 1970/71. Thereafter decennial censuses and in 1973/74 annual Agricultural Production Surveys were started.

2. Sources of data

Main data collection is done through an annual Agricultural Production Surveys (APS), decennial Agriculture Censuses (AC) and secondary data from other government departments. Both AC and APS's reference period is the 1st of August to 31st July the following year and this period is considered as the Agriculture Year for Lesotho. During APS the following information is collected at rural household level excluding urban area due to constraints of funding.

2.1 **Crop and Area:** collects information on area planted during two seasons (winter and summer), area harvested and production. Area is estimated through physical measurement of fields, which are operated by randomly selected households and production is estimated through sample-plots cutting. Information on fertilizers, pesticides, seeds, methods of ploughing, planting and weeding are estimated through interviews.

2.2 **Livestock**: Collect information during the first two weeks of August by interviewing households. Information covered is number of cattle, sheep, goats, horses, pigs and donkeys by age and sex, female cattle by main purpose, for which they are kept, livestock mortality and stock-changes.

2.3 In April, BOS and NEWU of DMA conduct Crop Forecasting Survey which is aimed at giving early estimates of crop production.

2.4 **Agriculture Price** is collected by Marketing Department of Ministry of Trade and Industry

4. Data gaps

Although BOS is able to produce some data on an annual basis there are still some gaps such as:

- Number of tractors used for agricultural production; Labour involved in agriculture annually; Prices at farm-gate and in the market; Imports and Exports of grain; Environment data such as: Soil types; Forestry; Production of Milk and eggs in urban areas.
- Farm management; Production of fruits and vegetables; Unofficial border trade of grain
- Production of cereal substitutes e.g Potatoes

5. Constraints/Problems

Agricultural statistics data is collected in the rural areas only due to lack of funds. Lack of co-ordination has been a problem, which has consumed a lot of the scarce resources. Duplication efforts could be eliminated through co-ordination among producers of data. There is an acute shortage of professional staff in a number of areas, especially data methodology and data processing. There is the problem of transport to enhance better and efficient fieldwork.

6. Suggestions/Recommendations

In order to consolidate the gains in improved data production in the field of agriculture, there is a need for technical assistance in the area of training, supply of more computers, vehicles and field equipment. Government needs to put very high priority on statistical development by increasing the budget for statistical production. In order to consolidate the gains in improved data production in the field of agriculture, there is a need for technical assistance in the area of training, supply of more computers, vehicles and field equipment.

References

M. E. MOLATOLI, 1996, Availability and improvement of agricultural statistics in Lesotho

FAO REGIONAL OFFICE FOR AFRICA, 1996, National User/Producer Workshop for Agricultural Statistics in Africa

O.O. AJAYI, 2000, Mission Report on Processing and Analysis of Agricultural Census Data (under FAO Project)

From Manure Administration to Manure Statistics

van Eerdt, Martha M.
Statistics Netherlands, Section Statistical Analyses, P.O. Box 4000, 2270 JM Voorburg, The Netherlands, e-mail: mert@cbs.nl

Abstract: Despite the disadvantages of registrations with regard to timeliness and quality, Statistics Netherlands increasingly uses administrations. The most important reason being cost reduction. Statistics on the use of manure and nutrients per municipality were produced from the integration of registered data from manure transfers with manure statistics. The statistical error of the results is quite acceptable. The results show that with the introduction of the nutrient accounting system (MINAS) the production and use of manure in the Netherlands further decreased. At the same time the transfers of manure to other farms also decreased leaving several regions with nitrogen surpluses.

Keywords: Animal manure, nutrients, balance sheet, statistics

1. Introduction

From the 1980's onwards the Dutch government developed legislation to restrict the pollution of soils, groundwater and surface waters by animal manure. The purpose of this legislation is to bring about an equilibrium between fertilisation and the withdrawal of nutrients by the crops. A certain loss is considered acceptable. In several regions in the Netherlands more nutrients are produced in animal manure than is acceptable. Farmers have two options to get rid of their surpluses: either reduce the amount of nutrients in the manure or bring the manure to other farmers.
The manure legislation requires farmers to registrate all manure transfers. These registrations are used in the Netherlands for the production of regional statistics on the use of nutrients in animal manure in relation to national standards and the standards of the EU Nitrate Directive.
The aim of this paper is to highlight some methodological problems and to present the results.

2. Material and methods

The use of animal manure per geographical unit is defined as the production in the region plus the input from other regions minus the output transferred out of the region. The smallest geographical unit is the municipality. The production is estimated on basis of the number of animals from the annual agricultural census and coefficients on the amount of manure and its content of nitrogen, phosphate and potash (van Eerdt, 1992; van Eerdt and Fong, 1998). The coefficients are updated annually.
From the registrations the amount of manure and nutrients imported to the region or exported from the region was deduced. Exports were also registered. Using existing

data from registrations has the obvious advantage that a survey is not necessary. A major disadvantage is the lack of control with regard to timeliness and quality of the data. The administrative data are gathered to charge levies. Statistics Netherlands harmonised definitions, made additional checks and surveyed the amount of manure and nutrients that was processed and purified. The calculated manure exported from the farm but not accounted for by transfer to another farm, a foreign country or a manure processing plant is considered to be stored or to be due to statistical errors. Part of the statistical error is due to remaining differences in definitions.
To integrate the production figures on nitrogen with the transport figures a correction was made for gaseous losses from stables and storage.

3. Results

At the national level the accuracy of the administrative data is quite good. Main discrepancies occur at community level. These inaccuracies at a regionally detailed level were overcome by aggregation of the manure types to total amounts of nitrogen, phosphate and potash.

Table 1: *Nitrogen and manure transfers in the Netherlands*

	Nitrogen (kg/ha)		Manure (ton/ha)	
	1997	1999	1997	1999
Nitrogen excretion	316	280		
Gaseous losses	64	51		
Manure	252	229	39,5	38,0
of which:				
Pasture manure	71	59	10,1	9,9
Manure put on the own land	125	109	21,3	20,3
Transfers to other farms	50	51	7,6	7,1
Exports	4	7	0,2	0,3
Processing and sanitation	1	3	0,3	0,6
Storage and statistical differences	0	0	0,0	0,1
Average manuring in the Netherlands	246	220	39,0	37,2

Average manuring in the Netherlands decreased by 26 kg of nitrogen per hectare between 1997 and 1999. Main reason was the decrease in production of animal manure. In 1999 over 20% of the manure produced on a farm was transported elsewhere. Most was transported to other farms. Of the manure produced 1% was exported and 2% was processed. Of the 20% of manure carried away from the farm where it was produced 12% is transported over short distances only. However long distance transports were relatively rich in nutrients.

Acknowledgements

Registrations on manure transfers were made available by Tom P.M. den Teuling of the Department of Levies of the Ministry of Agriculture, Nature and Fisheries. The author acknowledges his fine co-operation and helpful comments.

References

van Eerdt, M.M. (1992) Excretion factors of animal manure; calculation on the basis of statistics, in: *Work session on specific methodological issues in environment statistics,* ECE/Eurostat, Lisbon, 14-17 September 1992.

van Eerdt, M.M., Fong P.K.N. (1998) The monitoring of nitrogen surpluses from agriculture, *Environmental Pollution,* 102, S1, 227-233.

6 June 2001

Contributed Paper Session

THE AGRICULTURE-ENVIRONMENT RELATIONSHIP

Chair: M. Van den Broecke

Agri-Environmental Indicators: A Brief Overview on Data Need and Availability

Giampaola Bellini
Istat, Metodologie e statistiche ambientali, via A. Ravà 150, 00142 Roma, Italy
bellini@istat.it

Abstract: Agri-environmental topics are an issue which has only recently been highlighted and integrated into sectoral policies, starting in 1992 with the CAP reform and confirmed recently with Agenda 2000 adoption.
In general terms, decision makers and the general public need information for designing and monitoring an environmental policy. A system of statistical indicators can play an important role in this respect.
A brief overview on farm management indicators will be given.

Keywords: agri-environmental indicators, farm management practices

1. International organisms activities

Agriculture, as other human activities, can affect the environment in a negative way, locally and globally. In order to better describe the relationships established between human activities and the environment, the OECD first and EEA later worked on a conceptual framework, known as DPSIR (*Driving forces, Pressures, State, Impacts , Responses*).
Driving forces are an interesting and a not fully explored field of work in statistical terms. In the agricultural sector, farm management and agricultural practices can be classified as driving forces and their importance is well known since many political issues focus on them.

2. Data need and availability: an overview

International organisation are moving from indicator development to actual production of the proposed indicators. For this reason a check on data availability is necessary.
In this paper a brief overview on data need and availability at the national level is presented for *farm management indicators.*
At the beginning, the conceptual work has been concentrated more on pressure indicators than on driving forces indicators. Work on this issue has been done at Istat integrating new indicators (*italic text*) into the original OECD indicator list (OECD, 2001).

Indicators	**State**	**Note**
Environmental whole farm management plans	-	
Organic farming	-	
Soil cover *(see below)*	X	
Land management practices *(see below)*	X	
Nutrient management plans	X	
Soil tests	-	
Agricultural land under land management such as ***green manure*** *use (% of Utilised Agricultural Area - UAA)*	X	
Agricultural land under land management such as ***mulch*** *use (% of UAA)*	X	
Farms using ***crop residues burning practice*** *(%)*	X	
Agricultural ***land fertilised with organic manure*** *by kind of manure (% of UAA)*	X	Type of organic manure: *dung, slurry and liquid manure, compost*
Use of non-chemical pest control methods *Agricultural land, by kind of land under* ***Biological Crop Protection Method*** *(%)*	X	Type of land: arable land, permanent crop land, permanent grassland
Use of Integrated Pest Management *Agricultural land, by kind of land, under* ***IPM*** *(%)*	X	Type of land: arable land, permanent crop land, permanent grassland
Irrigation technology Irrigated land with different irrigation technologies applied (% of irrigated agricultural land)	X	Irrigation technologies: *flooding, flow through channels, aspersion, low pressure sprinklers, drip-emitters, other*

- : Not statistical data available
X: Statistical data available

Farm Structure Survey conducted by Istat in 1998 provided the information on agricultural practices mentioned above (Istat, 2001).

3. Conclusion

Farm management practices topic needs to be more deeply investigated for its political relevance, in fact many political issues focus on them. The Farm Structure Survey seems to be the main tool to collect data for agri-environmental indicators calculation. In fact questions are quite easy to answer by farmers and the information provided by them can be valuated in a better way by linking these data to farm structure data.
Nevertheless more efforts are still needed in order to better evaluate existing data and integrate environmental issues into sectoral statistics.

4. References

OECD, 2001. *Environmental Indicators for Agriculture – Volume 3: Methods and Results.* OECD Publication Service, Paris, France.

Istat, 2001. *Survey on the structure and productions of agricultural holdings: several agri-environmental aspects of the additional form for the year 1998*. Meeting of the Working Group "Statistics of the Environment". Joint Eurostat/OECD meeting. 19-20 february 2001 – Luxembourg.

Evaluating Agri-Environmental Measures of Rural Development Plans in Italy : A Set of Indicators[1]

Antonella Trisorio, Andrea Povellato, Annalisa Zezza
Istituto Nazionale di Economia Agraria, Via Barberini, 36 - Roma, Italy
e-mail: trisorio@inea.it

Abstract: The purpose of the research project is to design a methodological framework - common to all Italian Regions - for the evaluation of agri-environmental (AE) measures included in Rural Development Plans (Regulation EC n. 1257/99). Such methodological framework is intended to provide policy makers with a decision tool for properly evaluating the environmental effectiveness of AE measures. The evaluation studies on AE measures carried out in the past years showed the lack of a common analytical framework needed to better understand the environmental effects of AE measures, including cross-regional and temporal comparisons.
The formulation of AE indicators could provide useful tools to facilitate monitoring and evaluation of AE policies. Indicators are generally accepted as a tool for communicating summarised information about phenomena that are of relevance to decision makers. Such indicators are increasingly used for analysis and measurement of policy effects. Moreover, AE indicators would help to improve understanding of the complex relationship between agriculture and environment, to show its developments over time, and to provide quantitative information.

Keywords: policy evaluation; agri-environmental indicators; environmental effectiveness.

1. Research structure

The research project has been structured in three stages: 1) identification of potential environmental effects of AE measures; 2) identification of AE indicators; 3) selection of AE indicators.
Stage 1. The identification of potential environmental effects has been reached on the basis of an "actions/environmental objectives" matrix. The matrix was constructed according to the following two steps: a) identification of actions included in AE measures of the Rural Development Plans (RDPs) formulated by the Italian Regions. A comprehensive list of actions has been drawn up based on the analysis of the RDPs produced by Italian Regions. Actions have subsequently been grouped in order to facilitate the evaluation process; they represent the rows of the "actions/environmental objectives" matrix; b) identification of a list of "politically relevant" environmental objectives to be achieved through the implementation of the AE actions. The list of environmental objectives has been aggregated in five groups in order to facilitate the subsequent evaluation process; the environmental objectives represent the columns of

[1] The present research project is part of Inter-regional programme "Technical assistance, studies and researches" funded by Italian Ministry of Agricultural and Forestry Policies.

the matrix. The elements of the matrix show the potential contribution of each action as provided by the AE measures in meeting and/or conflicting with the environmental objectives identified.

Stage 2. The identification of AE indicators has been obtained starting from the potential environmental effects identified at stage 1, according to the following three steps: a) the potential environmental effects of listed actions have been identified according to five classes of intensity; b) only environmental effects pertaining to the two higher classes of intensity (high and very high) were taken into consideration; c) AE indicators were identified to measure environmental effects considered in the previous step. The identification of potential environmental effects and corresponding indicators have been worked out by environmental experts. The indicators identified refer to the "Driving forces - Pressure - State - Impact - Response" model. The latter is widely accepted at the international level, and particularly by EUROSTAT (European Commission). This model makes it possible to appropriately structure and organise the environmental information. In line with the specific aim of the project, attention has been focussed on pressure and state indicators. Indeed, pressure indicators make it possible: a) to better isolate the environmental effects of one specific policy from those generated by other policies and/or natural events; b) to detect phenomena quickly so as to give decision makers a chance to launch appropriate actions. Pressure indicators are generally able to signal any future environmental change before it actually occurs. Moreover, environmental objectives set in RDPs are often expressed in terms of changing pressures on the environment, based on assumptions about the relations between such variables and the environmental condition. The choice of state indicators was limited to such cases in which they were sensitive enough to record environmental changes within a “politically” acceptable lapse of time, that is in the short-medium term. The indicators were then combined with context indicators (referring to the features of the environment within the region) in order to better identify the environmental effect of AE actions. As a matter of fact, the same action generates different environmental effects according to the features of the environment on which it is exerted. AE indicators also need to be related to the farmland under agreement - broken down by type of zone - in order to better detect actual environmental effects.

Stage 3. The number of indicators should be limited enough to be manageable, but comprehensive enough to reflect the most relevant environmental issues. Indicators have therefore been selected on the basis of the following criteria: a) policy relevance in terms of addressing the key environmental issues faced by Regional governments; b) analytical soundness, that is they are based on sound theory; c) measurability in terms of availability of appropriate data in the short-medium term; d) ease of interpretation, meaning the providing simple and readily understandable information.

2. Results and concluding remarks

Results show that data availability represents a major factor in the selection of indicators. The most frequent problems were related to the level of aggregation (national vs. regional), the spatial scale (administrative vs. geographical), the lack of appropriate time series, temporal and spatial inconsistency, in addition to a more fundamental scarcity of environmental data. This state of affairs is mainly due to: a) the complexity and multi-faced nature of agri-environmental linkages; b) the relatively recent concern over agri-environmental issues.

Annex

ENVIRONMENTAL OBJECTIVES	TYPE OF ACTION	DPSIR	INDICATORS
Biodiversity conservation	ABCEGHILM	P	Area under agreement benefiting wild flora and fauna
	ABCEGHILM	S	Biodiversity of wild flora and fauna (presence, number, diversity)
	ABCDEG	P	Area under agreement contributing to diversity of intensively farmed agricultural habitats
	ABCDEG	S	Diversity of intensively farmed agricultural habitats (types of crops, expanse of fields, presence of ecotones, presence and diversity of crop rotations)
	EFGHL	P/S	Semi-natural or extensive habitats of agricultural ecosystems (number, area, mean area)
	GHILM	P/S	Uncultivated habitats of agricultural ecosystems (number, area, mean area)
	EGHL	S	Changes in crop varieties per species
	EGHL	S	Varieties of species and local and/or endangered varieties
	F	S	Changes in number of species and/or breeds under agreement
Landscape protection	EGHILM ACE	S	Visual coherence: structure and composition of landscape
	EGHILM AG	S	Coherence of flora, fauna and habitat
	EGHIM EGH	S	Coherence of landforms
	DHM	S	Coherence of water levels and contours
	EGHILM ACE	S	Landscape diversity (land use and lay out of crops)
	EFGHILM ACEGHI	S	Landscape diversity (flora, fauna e habitat composition)
	GHILM DFGHL	S	Landscape diversity (countryside stewardship)
	FGHILM CDFGH	S	Traditional crops/animals
	GI D	S	Linear elements
	GHIL DGM	S	Point elements
	ABCDEFGHIL	S	Traditional farming
Soil protection	E	P	Environmental restoration (conversion of arable crops in permanent meadows and pastures)
	CEGHI	P	Risk of soil erosion by water
	EFGHI	P	Risk of soil compaction
	E	P	Environmental restoration (set-aside of arable crops)
	E	P	Environmental restoration (crop rotations)
	CEHI	P	Tillage erosion
	EHI	P	Area with drainage systems and grass cover
	B	P	Mineral fertiliser use (NPK)
	BCE	P	Organic fertiliser use
	B	P	Methods and timing of fertilisation
	BC	P	Nutrient balance
	BCH	S	Organic matter content in agricultural topsoil
	A	S	Pesticide persistence in soil
Water protection	DEHI	S	Concentration of nitrogen in water
	DEHI	S	Concentration of phosphorous in water
	DEHI	S	Concentration of pesticide residues in water
	DEHI	P	Area subject to supported action reducing nitrogenous fertiliser
	DEHI	P	Area subject to supported action reducing phosphatic fertiliser
	DEHI	P	Area subject to supported action reducing pesticide use
	DEHI	P	Use of water for irrigation

Human health protection	A	P	Pesticide use
	A	S	Pesticide residues in food
Air quality protection	BF	P	Gross emissions of green-house gases from agriculture
	AG	P	Apparent consumption of Methyl-Bromyde
		P	Ammonia emissions from agriculture
	CEGHIL	P	Net absorption of carbon from land use changes and from forests

Legenda: A: pesticide use; B: fertiliser use; C: organic matter use; D: water use; E: cultivation of arable and permanent crops; F: livestock breeding; G: management of pasture lands and abandoned lands; H: landscape and natural resources conservation - whole fields; I: landscape and natural resources conservation - field margins; L: landscape and natural resources conservation - trees; M: landscape and natural resources conservation - isolated elements. DPSIR: conceptual framework; P: pressure indicator; S: state indicator.

Agricultural Resources Management and Sustainable Development

Igor E. Timchenko, Ekaterina M. Igumnova
Marine Hydrophysical Institute, 2, Kapitanskaya str., Sevastopol, 99011, Ukraine
E-mail: timchenko@stel.sebastopol.ua

Abstract: New agricultural resources management technology was suggested based on the Adaptive Balance of Causes (ABC) method. Examples of ABC technology implementation for decision-making support in various rural sustainable development problems are supplied.

Keywords: System management, development scenarios, new ABC-modelling, models of ecological-economics systems, data assimilation, soil, water and bio resources use, environment protection

1. The new Adaptive Balance of Causes (ABC) method of modelling.

The new approach to the complex systems management was suggested on the basis of three research results obtained by the authors:

1. System conceptions of sustainable development were formulated,
2. New Adaptive Balance of Causes (ABC) method for numerical modelling of complex systems has been worked out,
3. System management or general ABC-technologies of sustainable development were offered.

The new approach was constructed on non-linear causes - effect interactions (causes functions) operating inside a system (Timchenko et al., 1999). With the use of the Bernoulli type equation and special balance of causes condition, a standard local dynamic balance module was introduced. Causal functions and standard modules form the system's structure. Due to this, the ABC - method allows to transform a verbal expert description of the system into the prognostic numerical model and to obtain development scenarios (Timchenko et al., 2000).
Information ABC - technologies of sustainable development management were introduced, which permit identification of ABC - model coefficients on the basis of past and current data observations of the system state condition. Adaptation of development scenarios to the reality by data assimilation was suggested, based on optimal filtering technique (Timchenko, 1984).

2. System management of agricultural resources.

ABC - technology implementations to various economic, ecological-economic and socio-economic systems are considered. Examples of more than 15 dynamic systems of types: "firm - market", "market - enterprise - resources", "national economy", "Leontiev input-output balance", "bioresources consumption control", "environment pollution control", "production improvement paradox", "state investments distribution and social consciousness" are studied and discussed. Their prognostic development scenarios are obtained, and ABC - models reliability and flexibility are confirmed (Timchenko I. et al., 2000).

One of the most important practical task was to suggest a general management strategy for environmental bioresources use. It should be consistent with the problem of biodiversity conservation and environment protection against pollutant originated due to the conventional agricultural technologies. To solve this problem, the ABC-model of agricultural socio-ecological-economic system was developed. Rental for natural soil resources consumption and ecological fines for pollutant concentrations added to the environment were used as a main socio-economic management factor. General information technology of the decision-making in a rural socio-economic system was constructed, based on the simulated scenarios, generating by the ABC-model. Special adjustment procedure was proposed for development scenarios correction and the model adaptation to the observations coming from the suggested monitoring system.

General conclusions on system management and ABC - technologies efficiency for agricultural applications were made. The new system management approach has proved to be a very useful practical tool for decision-making support in various rural sustainable development problems: from private agricultural business planning to socio-oriented ecological-economic systems management.

References

Timchenko I.E., Igumnova E.M. Primalenny A.A. (1999). *Ecological-Economic Systems Management.,* “Ecosea”. Sevastopol. 180 p.

Timchenko I.E., E.M. Igumnova, Timchenko I.I. (2000). *System Management and ABC-technologies of Sustainable Development*, “Ecosea”. Sevastopol. 225 p.

Timchenko I.E. (1984). *Stochastic Modelling of Ocean Dynamics,* Harwood Acad. Publish. Chur-London-Paris-New-York. 1984. 320 p.

Collection and Use of Environmental Data

Milijana Ćeranić
Federal Statistical Office, Kneza Milosa 20, 11 000 Belgrade, Yugoslavia
e-mail: filippi@szs.sv.gov.yu

Abstract: Federal Statistical Office collects, processes and publishes, within its regular research program, numerous data from the field of environment. Considerable deficiencies in the functioning of environmental statistics are: uneducated staff, shortage of financial means and nonexistence of common data base.

Keywords:Environmental statistics, environmental protection, statistical investigations.

Environmental statistics in Yugoslavia is relatively young. It was developed together with the development of the unique statistical system of Yugoslavia, not as a separate statistical branch, but within statistical investigations in other fields, containing certain data on environment. Consequently, it was organised in line with that fact. For a short period of time environmental statistics functioned as an independent statistics, however, within reorganisation of the Federal Statistical Office, it was attached to the Statistics of Agriculture, Forestry and Environmental Protection.

Statistical data in this very important, complex and multidisciplinary area are also collected by some authorized institutions and organizations on republic and federal level, of which particularly important for statistics are Weather Bureaus and Public Health Institutes. Thanks to that, the Statistical Office has created a corresponding statistical data base containing data on quality of air, water, chemical composition of precipitation, rivers, rivers by river basins, lakes, water supply, water sources, public water supply, released water, use of water, irrigation systems, protection of water from pollution, waste and purified waters, damages caused by natural catastrophes, consumption of mineral fertilizers, consumption of chemicals for protection of plants.

In Statistics the following investigations are carried out: Report on Protection of Air from Industrial Pollution, Report on Use and Protection of Waters from Industrial Pollution, Report on Pollution of Air Across State Line on Long Distances, Report on Water Accumulations, Report on Public Water Supply, Report on Public Sewerage, Register of Protected Natural Entities and Individual Natural Monuments.

In addition to these investigations, Environmental Statistics uses the data from other statistical investigations processed in the Federal Statistical Office, as well as from other institutions dealing with such problems. It is difficult to say nowadays whether there exists a single statistical investigation which doesn't deal at least to some extent with the problem of environmental protection, since the number of inhabitants, living conditions, national income, the number of functional or dysfunctional vehicles, driving on good or bad roads, using regular or lead-free petrol, use of pesticides in farming, fruit-growing and forestry, uncontrolled cutting of wood, inadequate use of cultivable land, work of industry, discharge of litter and other, are all closely connected with mankind, that is, the environment.

At the moment, the statistical investigations mentioned are carried out with difficulties, because it still isn't possible to compel the reporting units to submit data to authorized bodies correctly and very precisely. The data are obtained on the basis of state laws and regulations, but mostly through interventions of statisticians in the field, as well as the enthusiasts in certain enterprises. In a small number of enterprises, predominantly the gigantic ones, there are departments with adequate staff who monitor all the relevant phenomena of importance for environment. Unfortunately, the number of reporting units is very small, therefore the resulting picture of environment is distorted. Besides shortage of staff and lack of awareness of the importance of environmental protection, it is obvious from our data that there exists a considerable shortage of financial resources, because numerous enterprises possess old and dysfunctional equipment, both as regards technological process, and equipment used for purification. In addition to that, the shortage of financial resources results in absence of measurements which, as a rule, are very expensive, or the measurements are carried out very sporadically, so that the obtained data are not representative enough to be published.

It is necessary to set up a unique data base of all data concerning human environment and establish a better coordination among all institutions which collect data, in order to avoid a situation in which some data would be collected by various institutions and the other data would not be collected by any institution. It is necessary to inform the public where what kind of work is done and published, so that the system would function in a unique manner. In order to assess as accurately as possible the environmental situation, we present the collected data on the level of municipalities, regions, provinces, republics and the state, and by waterways and river basins.

All data are joined and published in a regular annual publication titled "Ekobilten". The content of the bulletin is designed so as to cover all substantial fields of environment (a total of 22 statistical fields out of the existing 32): In the first place is a man, as one of the creators of environment in one hand, and degradation of environment in the other. It also covers other Eco-systems such as: water, air and soil from the angle of utilization, pollution and protection. Data on social product, investments, employment, education, transport, tourism etc. are presented as well as indicators of discord between the development of society and improvement of environmental protection.

The Available Data for Monitoring and Modeling Sustainable Development in the Agricultural Sector in the Republic of Serbia

Milorad Filipović
Faculty of Economics, Kamenička 6, YU – 11000 Belgrade
Email: miloradf@EUnet.yu

Vlade Zarić
Economics Institute, Srpskih Vladara 16, YU – 11000 Belgrade
Email: vzaric@EUnet.yu

Abstract: Agriculture in Serbia in one of the most important sector of its national economy. Together with related economic sectors, the agri-food sector contributes almost one third to the GDP formation. However, the structure of farms is very different – there are big state and small private farms. It is estimated that the important part of the agricultural production on the private farms is created in a the environment friendly way. Still there is no statistical evidence about organic production in the whole agri-food sector. Therefore there is a need for development for indicators of sustainable development, eco-statistic and green reports in the Serbian Agriculture.

Keywords: Agriculture in Serbia, Organic production, Indicators of sustainable development

In comparison with all other European countries where the importance of agriculture during the last decades has decreased, in Serbia has increased. In 1998, the contribution of agriculture to the gross domestic product was 21,6 percent (SGJ, 99, p. 129). The figure is much higher, however, when related economic sectors are also taken into account. Serbian agricultural export and import are important for the economy. The most important trade partner is EU and eastern European Countries, despite the sanctions imposed by EU and UN to FRY. The number of private farms, with an arable area less than 10 ha, increased from 1 092 762 in 1960 to 1 115 663 in 1999 (table 1).

Table 1: *Private farms according to size*

Area	Number of Farms, 1960	in %	Number of Farms, 1999	in %
< 0.50 ha	88 299	8.08	167 746	15.04
0.51 - 3.00	404 783	37.05	500 469	44.93
3.01 - 5.00	247 778	22.67	200 453	17.97
5.01 - 8.00	201 931	18.48	143 901	12.90
8.01 – 10.00	72 695	6.65	52 168	4.68
> 10.00 ha	77 276	7.07	49 926	4.48
Total	1 092 762	100.00	1 115 663	100.00

Source: Poljoprivreda Srbije 1947-1996, p. 397

The state owns farms are much bigger than the private, are organized as enterprises and co-operatives. The structure of agricultural area in agricultural enterprises, co-operatives and in private farms is shown in the table 2.
State Farms are basically not interested in the organic production and sustainable development. Since private farms are inappropriate for the "modern", i. e. intensive farming, they mostly produce in the environment friendly way. However, it is difficult to estimate the percent of the organic production of the total agricultural area and of the total number of farms. The statistics of the state authority shows that since 1992 the consumption of mineral fertilisers and plant protection preparation is declining.. It could be concluded that the whole production in Serbian Agriculture is made in the environment friendly way. However the decrease in consumption of mineral fertilizers was due to the high prices of the agricultural inputs in comparison to the agricultural output and misleading policy of protection of living standard by agricultural products price control, i.e. prices "scissors" had negative effects on agricultural production.

Table 2: *Agricultural area by categories of use and by ownership 1996 (thousand ha)*

Land by categories of use	Ownership			Total
	Enterprises	Cooperatives	Private farms	
1 Cultivable area	591 418	99 424	3 944 721	4 635 563
Arable Fields and Gardens	543 706	86 440	3 015 662	3 645 808
Orchards, Vineyards, Meadows	47 712	12 984	929 059	989 755
2 Pastures	117 888	33 328	518 866	670 082
3 Pools, reed tracts and fish-ponds	22 982	1 698	3 575	28 255
Total agricultural area)	732 288	134 450	4 467 162	533 900

Source: SGJ 1999, p. 221; SGS 1999, p. 186; Poljoprivreda Srbije 1947-1996, p. 51ff. and phone information by the Statistical Office of Serbia

In the 2000 the Serbian government introduced the law on organic agriculture (Sl. list SRJ, 28/2000) and now is too a short time to see its impact on the agricultural production. In short, in the Republic Serbia there are some initiatives for the sustainable development in the Agriculture, but only as initiatives and without data in the official accounting system as evidence. **Main findings are, first, there is need for statistical evidence in private farms, and second, for both private and state farms as well as for the whole agri-food industry, we need the indicators of sustainable development, eco-statistics. Green reports should be introduced in order to monitor sustainable development.**

References:

(1998) Poljoprivreda Srbije 1947-1996, 51, 397.
(1999) SGJ 1999, 221.
(1999) SGS 1999, 183, 186.
(2000) Law on ogranic agriculture, Sl. list SRJ 28/2000, 24-31.

Participatory Approaches in Information Gathering for Agriculture and Environmental Amelioration

Obot E. Ekop
Merbot Consult International Limited
P.O. Box 36150, Agodi P.O. Ibadan, Nigeria
e-mail: mailbox@mail-2.skannet.com

Abstract: The Nigerian policy on agriculture covers four components as applied in the Development plans, 1960-70, 1970-85, 1985-2000. These are Crops, Livestock, Fishery and Forestry. This paper shall seek to examine the information flow and management within this sector that would bring about sustainable environment.

Key Words: Agriculture, Environment, information, amelioration.

1. Introduction

An important indicator of socio-economic development is the quality of the environment in which the people operate. The link between environmental conditions and development are obvious. The word environment is multi-disciplinary as every sphere of human endeavour has its operational environment, political, economic, religious, agriculture, labour and others. These do not operate in isolation. Perhaps these must have informed the obvious neglect of a law to regulate our environment. Even when we talk of environment, the immediate thought is on water, air and industrial pollution, urban decay, waste disposal and lightly on deforestation. This is why the first Nigerian law on environment was Decree No 42 in 1988 which was promulgated in response to the Koko, Warri toxic waste episode involving the importation and dumping of hazardous waste in Nigeria. This was followed by the promulgation of the Federal Environmental Protection Agency (FEPA) Decree (Decree No 58) in December 1988. This charged FEPA with the responsibility of formulating and implementing a national environmental policy, among others. Estimates made in 1990 indicate that long term losses for only the top eight environmental problems in Nigeria will be over $5 billion annually if mitigative measures are not taken (Socio-Economic Profile of Nigeria 1996).

In a move to strengthen and give teeth to the activities of FEPA, Federal Ministry of Environment was created in 1999 to absorb FEPA's duties and responsible. "Its agenda being Environmental Renewal and Development Initiative (ERDI). The objectives are to take full inventory of the nation's natural resources, assess the level of environmental damage, design and implement restoration and rejuvenation measures and to evolve and implement additional measures to halt further degradation of our environment" (The Punch pp. 29 July 27, 2000). These are embracing a mission to cover the agricultural environment and its statistics.

The need for environmentally related statistics on Agriculture. Like other branches of official statistics, environmentally related statistics on agriculture serve the purpose of identifying problems, planning, execution, monitoring and all other public policy purposes. Agriculture is practised on the physical environment. Its practice is so crucial to the life of a nation that it affects the well-being of the people and the industrial development of the Nation. Its physical environment is static except reclaimed in terms of swamps or artificially created in terms of fish ponds. As the population grows, there is need also for the agricultural production to grow to meet the food situation of the people and as a source of raw materials for the industrial advancement of the nation. A sustainable physical environment is a pre-requisite requirement for sustainable agriculture. The lack of or the inadequacy of data for the design and management of the environment has led to the global concern for environmental problems confronting various nations.

2. Agricultural Practices

Agronomic practices for crop production, livestock, forestry and fisheries are carried out with harmful effect on the vegetation cover and environment, removing the protective role trees play for the sustainable functioning of the earth's life support systems. Hence desertification, loss of soil fertility, soil erosion, coastal erosion ground water resources pollution, declining fisheries, aquifer depletion leading to irreversible compactions, rural household time and municipal cost of providing safe water, loss of biodiversity, effects on regimes and ambient temperature, and many others.

3. Participatory Approaches.

Agriculture is a multidisciplinary subject with intra and inter-related disciplines. So also they are practised by many different professionals at intra and inter related fields. In Nigeria, these disciplines and fields are catered for by different Ministries and Parastatals. Moreover, the country runs a three tier government, the Federal, State and Local Government. Environmental Statistics as a new discipline needs to first evolve a concept and definitions with reference to agriculture. Then standardized indicators for measuring agricultural environment should be evolved in conference with the policy, research, practitioners-informal peasant operators. The environment is overburdened because the operators do not know what it takes to overburden the environment. When they are part of the formulation of the indicators they will know what it takes to SPOIL the environment and what information to supply per indicator and conserve the environment.

4. Conclusion

Statistics do not lie, but only if the correct information is given at the right time. To ameliorate the present deteriorating agricultural environment, the statistics to be generated need the involvement of all practitioners of agriculture and related disciplines in concept and in practices.

Gender Analysis: A Sustainable Path to Food Security

Mercy O. Ekop
Family Unity for Sustainable Development Foundation
P.O. Box 36150, Agodi P.O. Nigeria Tel: 02-2412626 Ext. 2390
Fax: 234-02-2411768
e-mail: library@odekulib.skannet.com - root@edulib.skannet.com

Abstract: The tremendous importance of gender role in food system to sustainably improve food security and eradicate poverty in developing countries is often overlooked. The gender analysis in this paper will help give this approach to food security appropriate consideration in future deliberations about the global food system at national and international levels. Hence, improve efforts to ensure sustainable access to sufficient food and income for all people.

1. Introduction

Ending hunger is not only more urgent than before, but is to be achieved more easily if individuals and government give it the priority it deserves. Most causes of persistent poverty and food security among rural women and the families they support are inter-related; lack of access and control of productive resources and services; over-and underemployment inequalities in employment opportunities and remuneration; exclusion from decision and policy-making; an unfavourable legal environment FAO (1996 - 2001). These causes should be addressed as priority areas to enhance food security, the status of rural women and achieve food security, poverty alleviation and sustainable development.

2. Gender, Food Security and the Environment

The nature and impact of gender on resources management and development strategies has either been overlooked or misunderstood by researchers and practitioners. And when gender consideration has been incorporated into natural resource development work, they have been approached sectionally with emphasis on the differences between men and women which must be accounted for in programme and project planning.
Throughout the world, both men and women are heavily involved in natural resource management and use. Their influence on the environment is outspoken in natural resource management and use. The influence on environment can differ greatly because of role differences. Men are often more involved in clearing fields, hunting and tilling the soil, while women frequently do more of the fuel and water collecting, cooking and weeding. By virtue of their task, their influence on and role in environmental management varies (table 1)

Table 1: *% Distribution of Productive and reproductive activities by men and women*

Activities	LGAs	Zaria		Katuru		Giwa	
	Gender	Men	Women	Men	Women	Men	Women
H/H Productive Life							
Crop		78	22	82	18	47	53
Livestock		30	70	48	52	26	14
H/H Reproductive Life		46	54	44	56	49	51
Average		51.3	48.7	58	42	40.6	59.3

Source: Ekop M.O. (2000) Field Survey

In Table 1, it is observed that in Zaria and Katuru men spent more time in productive life cropping for feed production while women spend more time in livestock farming than men. In reproductive life, there is a slight disparity between men and women ranging from 2% in Giwa to 12% in Katuru showing the percentage workload women carry over men in reproductive activities. In Giwa, strikingly, women spend more time and carry more work load than men in both productive and reproductive activities having a peak in livestock farming with 70% and 26% for women and men respectively. The advances of women are central to achieving sustainable development. Women are responsible for natural resource management though their day-to-day productive and reproductive tasks of providing fuel, water and food for household consumption and for sale. However, they are rarely in a position to influence decisions that determine the allocation of resources and hence, the sustainable use of land, water and woodlands, women have learned ecologically sustainable methods of agriculture and have acquired extensive knowledge about genetic diversity. If they are denied partnership in development, this wisdom is lost.

Women are often the major suppliers of household subsistence. When their access to productive resources declines, more people suffer from poverty and its related effects, including hunger, malnutrition and illness. Improving women's access to resources and services increases form productivity, provides a more efficient use of resources, and ultimately yields higher profitability, as women play a central role in the development of sustainable agricultural systems, particularly in improving crop and grassland. They play key roles in raising animals and in harvesting and processing livestock products both for home consumption and for sale. Although men are often the owners (and sellers) of large livestock, it is the women who perform most of the household labour devoted to animals. As men seek off-farm employment, rural women are assuming greater responsibilities.

3. The Roles of Gender in Developing Food Systems

The gender and development approaches seek to understand the ways gender constraints or advance efforts to promote while ensuring an equitable distribution of its benefits. Emphasis is placed on analyzing the incentives and constraints under which men and women work in order to make visible the differences between them in terms of their roles and work loads, the impact of intervention on them, and their ability to gain access to resources and decision making. It also analyses the implication of these findings for planning and implementation.

Efficiency, a central component of approach, provides the tools and methods for more sound intervention. Greater participation is sought from women at all levels. Women play a central role in the development of sustainable agricultural systems, particularly in improving crop and grassland, productivity. In sub-saharan Africa, it is estimated that women contribute 30 to 80 per cent of the agricultural labour for crop production, depending on area and economic class.

Rural women are producers of food, traders, and family care-takers. They play an important role in their communities as well as in national economies. Their efforts to initiate or expand income generating activities, however, are constrained by their limited access to credit and other financial services such as savings and deposits. Access to the services would ensure sustainable financial intermediation and discourage dependence on external sources.

Financial services must be made available if small farmers are to improve their agricultural productivity and enhance their house hold income and food supplies. They need credit to purchase Agricultural inputs such as improved seeds, fertilizers, insecticide and herbicides or to higher paid labour etc. Women lack of access to credit, which is part of a larger problem of inadequate credit availability for small farmers.

Women as well as men, need increased access to appropriate financial services such as savings, deposit and credit. They also need a greater capacity to negotiate with formal rural finance structure as well as information and technologies and more varied roles in managing the family farm, including animal husbandry operations. Rural women are major caretakers and users of forests. They are main gatherers of fodder and fuelwood, and they seek out fruits and nuts to provide for their families. In addition, they use roots and herbs for medicines. Women's gathering activities are very important to household income and nutrition. The products they gather are processed or marketed, bringing in supplementary cash income. Women contribution to fisheries is substantial. In some regions women are engaged directly in fish production. It is then necessary to ensure that women are equal partners and productive and self-reliant participants in fisheries activities aimed at improving their own and their families nutritional and living standards.

4. Conclusion

It is important for development experts to feel compelled to think about men and women independently, primarily because independent consideration increases the potential or design, implementation and management of effective, sustainable development activities. Decision-making process, should be participatory if sustainable food security is to be achieved.

References

Ekop M.O. (1999). Gender Implication in Project Design. A way formed in development Process. Paper presented at the International Biometrics Society Annual Conference 1999 at IITA, Nigeria.

Ekop M.O. (2000). Gender consideration on market-oriented smallholder peri-urban dairying in the inland valleys of West Africa (A case study of Nigerian situation). A survey conducted for International Livestock Research Institute (ILRI).

Gender key to sustainability and food security EAO - Plan of Action for women in Development 1996-2001

Environmental Statistical and Mathematical Analysis of Pesticide Applications in Armenia

Vardan Sargsyan
Yerevan State Institute of Economy, Nalbandyan 128, Yerevan, 375025, ARMENIA
e-mail: vardan@ysine.am

1. Introduction

Environmental problems connected with agriculture are common, not only for developing countries. In order to protect nature and the health of people, it is possible to refrain from nuclear technologies or reduce and close some such productions. The same can't be said about agriculture, where chemical substances dangerous for nature and people's health are nowadays widely used. Though nowadays the scientists offer non-chemical means of agricultural production, nevertheless, none of those methods can substitute completely for pesticides. The only way out seems to be adopting a "favorable" strategy of pesticide usage in agriculture with appropriate risk assessment to achieve health safety and optimal environmental quality. This paper presents the problem of pesticide applications in Armenia, a description of research realized concerning the questions of risk assessment and optimal control of pollution and directions of future investigations.

2. The Conditions and Problems of Pesticide Applications in Armenia

Agriculture is one of the main branches of economy of Armenia, where industry generally is not working yet, after the breakup of the USSR. Agricultural production here is unable to meet the demand and several important agricultural products are imported. Due to limited land resources (1.3 million hectares) application of intensive methods of plant cultivation are in force. Armenia is one of those NIS countries, where, especially during soviet period, pesticides were used in inadmissible high quantities. If in the USSR in general was used 0.5 kg. pesticides per ha, in Armenia it reached 10-15 kg/ha, and in some regions - up to 35 kg/ha.

In two main agricultural regions (the Ararat Valley and the Gegharcunik District) intensive pesticide application had generated environmental and health problems. Risk-benefit analysis of pesticide applications is the basis of developing sustainable agricultural policies is those regions.

The environmental problems of pesticides in the two above mentioned regions of Armenia present typical examples of negative influence: increase of morbidity and loss of environmental quality, particularly contamination of the freshwater basin. The questions of risk assessment and optimal policy choice are of high importance for the integration of environmental sustainability and economical growth. With the help of economical modeling tools investigations are carried out in two directions:

- Working out an optimal strategy of pesticide usage on the basis of taking into account ecological damage,

- Optimal policy choice for pollution regulation and obtaining an optimal environmental balance.

3. Policy Choice Problem

Investigations into the problem presented here will be inadequate without a policy choice as environmental degradation arises as an economic policy problem. The government has a number of instruments available for pursuing policies aimed at improving environmental quality. The main choice is between Pigovian taxes and quantity regulations.

The latest achievements in environmental management theory can be applied to investigating this problem in the Ararat Valley and the Gegharqunic District to evaluate alternative policy instruments in order to achieve environmental improvements associated with pesticide pollution control. Methods of a comparatively new nature developed in Environmental Economics-Weitzman's theory of control under asymmetric information [6] are more acceptable for the economic situation in Armenia. The problem arises from differences in the amount of information possessed by the polluter and the regulator. Most frequently, this difference is that polluter has private information that the regulator needs. These methods are suitable under conditions of present agricultural production in Armenia as the land is already private. The newest achievements in this theory form the basis for development of the research. Recently several investigations have extended this analysis to dynamic setting, where the externality is caused by a stock rather than a flow, particularly the model by M.Hoel and L.Karp "Taxes Versus Quotas for a Stock Pollutant" [2] and its developments. The model defines acceptable tools and optimal control rules for environmental regulation in a dynamic context on the basis of Benefit and Damage functions. The application of this model for the problem presented above generates scientific interest and is a way of working out a sustainable pesticide management policy in Armenia.

4. Concluding Remarks

Countries in transition have inherited a badly damaged environment, energy and raw material-intensive economies, and obsolete, polluting technologies. Furthermore, they are facing new constraints associated with the shift from a centrally planned to a market economy. However, it is widely understood that the aim of transition is not only to improve economic performance in the short and the long term but also to move towards sustainable development.

The results obtained show the necessity of carrying out further investigations of the problem. It should also provide answers to questions which are of scientific interest:

- How has the transition to a market economy and the attendant economic reforms in Armenia affected environmental quality?
- How do transition conditions affect the choice between environmental taxes and tradable pollution permits in Armenia?

The examples of environmental situation around Ararat Valley and Lake Sevan are very common in the developing world, where environment suffers very often from economical needs and where constrained economical policy brings numerous environmental problems.

6 June 2001

Contributed Paper Session

DATA PROCESSING

Some Experiences in Combining Different Methodologies for Editing Agricultural Data

Marco Di Zio
ISTAT, Servizio MPS, Via Depretis 74/b 00166 Roma, Italy e-mail:dizio@istat.it

Orietta Luzi
ISTAT, Servizio MPS, Via Depretis 74/b 00166 Roma, Italy e-mail:luzi@istat.it

Abstract: The main objective of the paper is to underline the complexity of building an integrated editing and imputation process where different methodologies are all combined together to deal with the different types of errors affecting agricultural data. To select the '*best set of solutions'* for classes of agricultural surveys, experimental studies are needed. Two experiences performed in this area are presented.

Keywords: selective editing, probabilistic editing, imputation

1. Editing agricultural businesses

Designing an editing and imputation (E&I) strategy for data collected on economical subjects, like agricultural firms, is a complex task because of a number of reasons: the nature of phenomena investigated, the characteristics of surveyed units, the survey objectives, the particular kinds of non-sampling errors. Because of the complexity and agricultural data peculiarities, the design of an E&I strategy requires in these cases an accurate analysis of the most appropriate approaches for dealing with the different kinds of units/errors/information characterising these surveys. Hence a preliminary phase is needed where *hierarchies* are defined between: firms (more or less relevant in terms of dimensional or economical criteria); variables (target, other variables); errors (systematic, influential, stochastic, etc.). This is equivalent to splitting the overall E&I problem into several sub-problems, each dealing with a particular kind of firms and/or variables and/or errors. For each sub-problem, the need of identifying the corresponding *best solution* among candidate methodologies in the particular survey context is the crucial point. For this aim, experiments are needed for testing the efficiency of alternative techniques and optimising each particular E&I step. The concepts of *efficiency* and *optimisation* depend on different aspects: the survey objectives (e.g. accurate and coherent estimates and/or coherent micro data); the availability of auxiliary information (e.g. external sources or historical information); available resources (e.g. budget, time). In addition to those elements, other ingredients are important for the choice of the *best solution*, like the nature of collected variables (e.g. structural and/or economical, collected at aggregate or detailed level, etc.), and the questionnaire structure.

2. Some examples of combined techniques in the agricultural context

In this section two integrated E&I procedures for agricultural surveys are synthetically illustrated: the experimental procedure for the 1997 pilot survey on Balance Sheets of

Agricultural Firms (BSAF) (Di Zio & Luzi, 2000a) and the test procedure carried out in the context of the design of the new E&I strategy for the 2000 Italian Agricultural Census (IAC) (Di Zio & Luzi, 2000b). The BSAF sample survey collects some economic information (e.g. costs, stocks) on 2067 firms. In this survey we worked in the situation where poor statistical/mathematical relations between variables exist. This depends on both the questionnaire structure and the peculiar economical characteristics of the agricultural firms surveyed. On the other hand, the IAC collects both structural information (e.g. kind of management) and production variables (e.g. surfaces, specific growing) strictly linked between themselves. In both studies, the goal was verifying the usefulness and the feasibility of a new E&I approach. In particular, in the case of the IAC, the proposed E&I strategy has been compared with the entirely deterministic 1990 procedure. Both proposed procedures consist of the following main phases: 1) *Pre-editing*: detection and treatment of systematic errors; 2) *Selective editing*: detection and review of relevant or influential errors; 3) *Automatic E&I*: detection and imputation of non-influential errors. In both strategies, the methodologies tested in each phase were: an approach based on the use of *deterministic rules* in the pre-editing step; *selective editing functions* (Latouche & Berthelot, 1992) to identify influential errors; the probabilistic *Fellegi-Holt methodology* (Fellegi & Holt, 1976) (FH in the following) in the error localisation step and a *hot-deck nearest neighbour donor* method in the data imputation step of the automatic phase. In the two cases different selective functions have been used to take into account the particular characteristics of the two surveys (census/sample), the different nature of the variables investigated, the available resources and time (few in BASF, more in IAC). In the two applications very different performances have been produced by the probabilistic error localisation algorithm, because of the different structure of questionnaires and the relations between variables. It is well known that the FH approach for detecting errors, based on the minimum change criterion, gives to each variable (involved in the set of edits) a probability with an error proportional to the number of failed edits involving that variable. The consequence is that in the case of few and non-strictly connected edits, the algorithm can be too much "free" to localise the error: in this context a deterministic approach could perform better. This happened in the BSAF survey, where generally only one edit can be defined for each variable. On the contrary, in the IAC, where each variable is involved in more than one rule, the FH algorithm performed very well at both micro and aggregate level with respect to the traditional entirely deterministic E&I procedure.

References

Di Zio M. & Luzi O. (2000a) Comparing a purely deterministic and a semi-probabilistic approaches for the E&I of Agricultural Census data, *Internal Report*

Di Zio M. & Luzi O. (2000b) Combining methodologies in a data editing procedure: an experiment on the survey of Balance Sheets of Agricultural firms, *Internal Report*

Fellegi I.P. & Holt D. (1976) A systematic approach to edit and imputation, *Journal of the American Statistical Association*, 71, 17-35.

Latouche M. & Berthelot J.M. (1992) Use of Score Function to Prioritize and Limit Recontacts in Editing Business Surveys, *Journal of Official Statistics*, 8, 3, Part II.

The GIS and Agricultural Data

Slavko Kapuran, Ema Jovanovic
Agricultural Division, Republic Statistical Office of Serbia
11000 Belgrade, Milana Rakica 5, tel +38111411836, fax +38111411260
e-mail: filippi@szs.sv.gov.yu

Abstract: This paper presents some experiences in using GIS in the field of agriculture data of the Republic Statistical Office of Serbia. A data base of geographical data defining administrative and cadastral communes is provided by the Republic Geodetic Office. The area information component is integrated with item description and data collected by agricultural statistics, so as to analyse data quality in the phase of data collection (done by subjective method of estimates) as well as analyses and presentation of data on the level of communes. GIS output provides a visual presentation in a graphical context and makes phenomena more understandable and attractive for end-use.

Keywords : GIS, data presentation, method of subjective estimates.

1. Introduction

In 1996, Republic Geodetic Office (RGO) completed geocoding borders of administrative communes (189) of Serbia, which was a pre-condition for introducing the Geographical Information System technology (GIS) in official statistics. As a software tool for creating thematic maps and area queries for analysing data, the Republic Statistical Office (RSO) introduced Intergraph's GeoMedia.
In 1997, after introducing experiences of Israel Statistics about using GIS in organising and conducting population census, it was decided, together with RGO, to investigate possibilities in applying similar solutions in the Census 2001. Unfortunately, this idea was left, due to lack of sources for preparation activities.
In 1998, RGO geocoded the whole network of cadastral communes. A cadastral commune is one of the basic area units of the geodetic information system, connected with other area units and item attributes that are in the competence of RGO. There is also a direct connection between cadastral communes and statistical area units (statistical and census areas) maintained by RSO. The concept of the cadastral commune is directly used in agricultural statistics.

2. Agricultural statistics

The share of agriculture in the GDP is about 22% and the share of agricultural products in total exports is about 25%. The agricultural sector employs 17% of the population.
Agricultural statistics investigate and provide yearly data about storage, production, trade and consumption of agricultural products. Due to specifics of agriculture as

organic production, as well as its agrarian and organisational structure in Yugoslavia, different statistical methods for collecting data are in use. Objectively, the report method, based on accountancy records is present in obtaining data about crop production and livestock breeding, from agriculture companies. Next to this, there are methods of subjective estimates used for gathering data about crop production on private agricultural households, through a network of estimate areas and statistical estimators. By sampling method we obtained data about livestock breeding in private agricultural households. Measured by physical volume of agricultural production, data obtained by the report method cover 18%, by method of subjective estimates 46%, and by sampling method 36%.

3. Method of subjective estimates and GIS

By this method, data about crop production of the privately owned sector, which covers approximately 85% of cultivable area (3.9 mil ha), are provided. Data are collected for 79 cultures and their status is recorded in 7 time points during their reproductive cycle.

A network of statistical estimators consists of 1450 persons, who are agricultural experts or agricultural producers, who estimate for selected plants, the area, average and total yields on the estimate area. Average cultivable area per estimator is 2800 ha. In Serbia there are 5132 estimate areas consisting of one or more cadastral communes (5826), each covering approximately 800 ha (average) of cultivable area.

On the basis of RGO data about category of land usage and ownership on the level of cadastral communes, RSO defines the structure of land of the privately owned sector for each estimate area. Estimators, by visiting and observing a certain area and by estimating on the basis of available data about seeds used, fuel, fertilisers estimate structure of sown area, average and total yields for their estimate areas.

The reliability of collected information, taking in consideration the large number of units (over 1.1 million), small size of holdings (3.5 ha of cultivable land) and lack of data on rural households, presents problems which are not solved in a satisfactory way, until today. GIS is used for testing and improving data quality obtained by the method of subjective estimates. By area queries, differences in average yields in neighbouring estimate areas, or by categories of land similar areas, can be analysed. There is also a possibility for testing rationality of statistical estimators network from the point of costs, burden of estimators (size and homogeneity of area) and their results. Certainly, more qualitative analyses could be done if the data about local climate characteristics, river flows and communication network were included in the data base.

4. GIS and data presentation

Since the estimate area is a part of the administrative commune, by summing the estimate area’s data it is possible to get the data about crop production of the privately owned sector, for communes and higher territorial levels. Since the participation of private owned sector in GDP of whole agriculture is 80% in Serbia (in some communes that participation is over 90%), interest for this kind of data is significant. GIS enables, using different area queries, analyses of crop production by communes, their significance for certain plants, creating thematic maps for publications and presentations for different users. In the data base there are 510 items for last three years (1998-2000), with 250 000 data.

5. Conclusion

This project realised in GIS manner, pointed out advantages in using agricultural data connected to the area component. Further development of the project requires more detailed analysis of the data base content and available data, needs of the end-users and term plan for sources for realisation, implementation and user training.

Economic and Environmental Analysis of the Italian Agricultural Sector: the ELBA Model ([1])

Giuseppe Palladino, Marco Setti
University of Bologna – Diproval – Economics Unit
Via Rosselli 107, 42100 Reggio Emilia, Italy
e-mail: palladino@stpa.unibo.it – msetti@stpa.unibo.it

Abstract. The ELBA Model (Environmental Liveliness and Blended Agriculture) is a partial equilibrium sector model focused on the analysis of the Italian agri-environmental systems at regional, provincial, and local level for Italy as a whole. Based on the PMP methodology, the ELBA Model has been developed in order to monitor, and to forecast the economic, political and environmental items related to 63 primary production activities. ELBA has been designed as a Web tool for the public administrations.

Keywords: model, agriculture, economics, environment, policy

1. The ELBA Model

The sector model ELBA (Environmental Liveliness and Blended Agriculture) is oriented towards the analysis of the Italian agri-environmental systems. It aims at promoting the sustainable development of agricultural activities. The ELBA Model depicts the overall interactions both between the production activities (47 crops and 16 animal categories), and between the latter ones and the natural media in order to monitor, and to forecast the economic, political and environmental factors related to the Italian agricultural sector.

ELBA is based upon a consistent multi - spatial (national, regional, provincial and local level) and multi – field (technical, economic, political, environmental, orographic, pedological, climatic variables) time series data set organised through a specific DBMS. As regards the production related part of the data base, an activity based accounting system consistent to the national Economic Accounts for Agriculture (EAA) has been adopted.

The basic data sources are Eurostat's REGIO, CRONOS and COMEXT, ISTAT, and FADN data bases. Lacking data on crop production, and livestock (e.g. feed uses, young animals flows) at the different spatial levels are endogenously derived. Interregional trade is derived as well. Furthermore, the data base includes soil data (source: MIPAF), land use (remote sensing - 1 km^2 grid), and weather time series (source: European Centre for Medium-Range Weather Forecasts).

ELBA reaches a comparative static (partial) equilibrium through an iterative process which couples the supply behaviour with a market component. The market module is a non–spatial multi–commodity one which returns market clearing prices. The supply side

[1] The present paper is financially supported by University of Bologna (funds for selected research topics).
Setti M.: Section 1.; Palladino G.: Section 2.

is characterised by a modular design where single sub-modules are interrelated and, in general, based on a wide series of engineering functions the outputs of which are consistent to the national EAA.

2. Results

Inside the supply module, the "herd sub-module" describes the demographic evolution of the single animal categories, defines the related interregional trade flows, and derives the yearly number (and final weights) of animals bred in each area. The "feed sub-module" quantifies analytical feed consumption and costs for each regional livestock. The "environmental sub-module" estimates the emission of organic nutrients as a balance at "animal level", and the surplus of (organic and inorganic) nutrients as a balance for each crop activity and area.

ELBA simultaneously solves non-linear optimisation problems (multi-activity cost functions for all the 103 Italian provinces) through the Positive Mathematical Programming and Cross Entropy methodologies. Technical and economic results are used to depict the production activities of the Italian province covering the whole agricultural national land. Through GIS overlay and spatial processes, weather conditions and soil profile are then taken into consideration, and bio – physical models (CGMs involving the phenological characteristics of each single crop) are integrated into the economic modules to describe the fate of the main nutrients for each cell in detail.

Hence, for each basic spatial unit in which the 103 Italian provinces are divided, ELBA links and measures the environmental performances of the agricultural activities, and quantifies the deriving emissions of the land, water and air pollutants. The surplus and the impact of nutrients cycle with global warming potential are described both for administrative districts (region, province, …), and for natural areas (catchment, vulnerable areas, single cells, …).

Results for economic (costs, gross margins, subsidies, …) and environmental parameters are obtained through reference (base period) and simulation runs (Agenda 2000 scenarios, cross-compliance) for Italy as a whole. The obtained results are directly comparable at regional level with the EU-wide CAPRI Model outputs. ELBA is furthermore designed and as a GIS-WEB tool available for the public administrations.

References

Heckelei T., Britz W. (2000) *Concept and Explorative Application of an EU-wide, Regional Agricultural Sector Model (CAPRI- Project)*, in Agricultural Sector Modelling and Policy Information Systems, edited by Heckelei T., Witzke H. P. and Henrichsmeyer W., 2001 - Wissenschaftsverlag Vauk Kiel KG.

Howitt R.E. (1995) "Positive Mathematical Programming", in *Amer. J. of Agr. Econ.*, 77, 329-342.

Palladino G. et Al. (1998) *A regionalized analysis of the environmental impact of the animal production activities - nitrogen and methane emissions*, Working paper 98-11, CAPRI research project, University of Bologna, Bologna.

Modelling Agricultural Experiments

Stanislaw Mejza
Agricultural University, Wojska Polskiego 28, PL-60-637 Poznan, Poland,
e-mail: smejza@owl.au.poznan.pl

Abstract: The purpose of the paper is to illustrate an objective procedure for derivation (definition) of the linear model of observations obtained in agricultural experiments. In our considerations the process of randomisation plays a central role.

Keywords: additivity, blocking systems, conceptual response, modelling, scheme of randomisation, factorial designs, designs with split units

1. Introduction

The experiment is an important tool of research in agricultural sciences. Hence, planning, modelling and inference problems are of fundamental importance for every experimenter using experiments in his research work.

There are, in practice, two main approaches to the model building of a linear model of observations. In the first approach we assume a priori a form of the linear model, usually before performing the experiment. The linear model and its dispersion structure are assumed to be independent from the type of experiment and on the structure of the experimental material. Sometimes, some additional assumptions concerning dispersion structure (correlation, auto correlation, etc.) are added. The problem is how to check these assumptions.

In the second approach, the model is strictly connected with a given experiment i.e., with the structure of its experimental material and with the method of assigning treatments to the units. This is given by the so-called scheme of randomisation.

At the beginning let us consider the factors which have an influence on the value of the observed data (called also observed response, observed yield).

Note that every unit (plot) possesses some kind of fertility which gives some yield in the case when treatments do not occur on a unit and in the case in which no treatments have an effect on the yield. This yield will be called zero yield (conceptual response). The increase (or decrease) in zero yield due to the treatment used on the experimental unit will be called pure effect (due to treatment). Usually the sum of zero yield (conceptual response) and pure effect due to treatment is called the pure yield (pure response) and is often the base of the statistical analysis.

Let us note that when observing the response of the unit in reality, any observation may be affected by a "technical error", an error due to some technical inaccuracy in performing the experiment and due to some error connected with measurements of the response (data). This error is also called measurement error (cf. Neyman et al., 1935).

In our approach, it is assumed that observed response is a sum of three components: a conceptual response connected with an experimental unit, a pure effect due to treatment

(combination) and a technical effect connected with measurements. It means that additivity among these three components is also assumed.
The paper deals with the problems connected with model building for popular types of designed experiments. The one- and two-factor experiments carried out in design with one or more blocking systems are taken into account only. A block design, a nested block design, a row-column design, and a block design with nested rows and columns are considered for one-factor experiments (cf. Mejza and Mejza 1989, 1994). For two-factor experiments a classic two- (many-) factorial design, a split-plot and split-block design are considered only (cf. Mejza, 1994). The former designs obey the incomplete, complete and over complete cases of these designs.
In this paper we present several different schemes of randomisation's for the designs mentioned above. Moreover, the so-called comparative experiments are taken into account only.
The starting point of our considerations is some theoretical (master) plan of the experiment. In this plan, say *D*; we take into account all the experimenter's suggestions concerning the statistical properties of design and the experimental conditions.
It means that plan *D* will not be chosen at random. The basic problem worked out here, is the way of assigning plan *D* to a given experimental material. This is defined by the scheme of randomisation. It describes how to assign the theoretical units of plan *D* (with their treatments), to the experimental plots. In our considerations the treatments will not be randomised. Suppose that the randomisation is performed as described by Nelder (1954) by randomly permuting, for example for block design, blocks within their total area and by randomly permuting units within blocks.
It will be assumed that the treatments under consideration are homogeneous (or additive) in the sense that the variation of the response among the available experimental units does not depend on the treatment received (cf. Kempthorne, 1952, Nelder, 1965).
Finally, a special attention is paid to validity, with respect to the randomisation point of view, of linear model assumptions adopted in many applications.

References

Kempthorne, O. (1952) *The Design and Analysis of Experiments*. Wiley, New York.
Mejza I., Mejza S. (1989) A note on model building in block designs. *Biuletyn Oceny Odmian*, 21-22, 87-201.
Mejza I., Mejza S. (1994) Model Building and Analysis for Block designs with Nested Rows and Columns. Biom. J. 36, 327-340.
Mejza S. (1994) On modelling of experiments in natural sciences. Listy Biom.- Biom Letters 31, 79-100.
Nelder J.N. (1954) The interpretation of negative components of variance. Biometrika 41, 544-548.
Nelder J.N. (1965) The analysis of experiment with orthogonal block structure. Proc. R. Soc. Lond. A. 283, 147-178.
Neyman, J., Iwaszkiewicz, K. and Kolodziejczyk St. (1935) Statistical problems in agricultural experimentation. J. Roy. Statist. Soc. Suppl. 2, 107-154.

Interest of Numerical Meteorological Model Parameters as Input in the JRC Crop Growth Monitoring System

JM Terres,
JRC SAI– Ispra TP 262 – EGEO unit, 21020 Ispra (VA), Italy
e-mail: jean-michel.terres@jrc.it

P. Segers, F. Biard,
GEOSYS, 20 Imp René Couzinet, BP 5815, Toulouse Cedex 5, France

R. Delécolle
INRA Bioclimatologie, 84914 Avignon Cedex, France

Abstract: The objective of the study is to test meteorological Global Circulation Model (GCM) data for being used instead of interpolated stations observations in the EC-JRC Crop Model for yield forecast. One year (1997) of weather parameters are provided by ECMWF and Meteo-France models. GCM data are first compared with MARS interpolated observations, then used as input in the crop simulation model.
Despite some differences in the precipitation field, consistent results demonstrate the potentiality of GCM for continental scale crops monitoring.

Keywords: agrometeorological modeling, global circulation model, crop monitoring.

1. Introduction and Background

The MARS project from the Joint Research Centre has been set up to provide the EC DG Agriculture with homogeneous and timely information on agricultural production. To achieve this goal, agrometeorological modeling techniques have been combined into a GIS to provide crop growth monitoring capabilities. The system is driven by meteorological conditions, modified by other environmental factors such as soil characteristics and crop parameters. The current system uses daily meteorological observations acquired by stations from the WMO network covering Europe, Turkey and North Africa. Unfortunately, meteorological observations are heterogeneous in time and space, therefore alternative data sources coming from meteorological Global Circulation Model (GCM) are tested. Their main advantages are data homogeneity and cost.

2. Methodology

Data sets for 1997 are provided by ECMWF (Reading, UK) and by Meteo-France (MF). GCM outputs are compared with the MARS interpolated meteorological observations on grid cells of 50*50 km (taken as reference in this study).

Comparison of the three meteorological data sets and comparison of corresponding simulated crop indicators for soft winter wheat (*Triticum aestivum*) are made. Assessed crop parameters consist of phenological development, Soil Moisture, Biomass (kg/ha) and weight of grain (kg/ha). Data are analysed using classical statistics (mean, standard deviation, percentiles) and graphical tools (Scatter-plots, Maps of differences, Plots of temporal evolution).

3. Data Analysis - Discussion

Temperature parameter is found with best consistency between GCM and MARS. However, GCM tend to underestimate this field, resulting in a delayed development stage during crop growth simulation.
For precipitation, both models have significant overestimation. Intra-annual analysis shows the overestimation occurring in autumn-winter for ECMWF and in summer for MF, mainly in southern areas. With the improvement in GCM surface scheme, one can expect an improvement in the precipitation distribution.
The overestimation observed for Global Radiation and Potential EvapoTranspiration may partly explain the slight biomass overestimation obtained with GCM.
Even if the final level of crop production is close between all data sets for 1997, additional tests should be carried out using longer time series to determine whether climatic accidents are well reproduced by GCM.

4. Conclusion

The study demonstrated that GCM outputs can be used as input in a crop monitoring system even if some parameters such as Rainfall or Global Radiation show discrepancies between tested GCM and observations data. Such implementation requires beforehand to re-calibrate the crop model for these alternative data sets. However, the rapid increase in GCM resolution and their constant gain in accuracy make them definitely of primary interest for crop monitoring at a continental scale.
Furthermore, the use of meteorological models opens new possibilities such as agri-environmental applications (i.e. pesticide and fertilisers fate modeling, soil erosion, natural hazards risks assessment) at a local scale through the use of a Local Area Model (LAM) or such as early crop yield assessment through the use of seasonal weather forecast (in region with clear predictability i.e. tropics).

References

Segers P., Terres J.M., Biard F., Orlandi S., Delecolle R., 1999. Interest of numerical meteorological model parameters as input in the JRC Crop Growth Monitoring System. Report EUR 19030 EN, 1999. 137p.

Terres J.M., 1999. The Crop Growth Monitoring System implemented by JRC/ARIS unit for the information needs of the EC DG VI – Agriculture. In: *Proceedings of the international symposium Modelling Cropping Systems – European Society for Agronomy. Division Agroclimatology and Agronomic modelling. EX 18746/C.*

Guarding Quality of Agricultural Statistics – An Indian Model

T. Baskaran
Joint Director, National Statistical Commission, Government of India, M-38/1, IBC Building, Middle Circle, Connaught Place, New Delhi 121001, India
e-mail: nsc@nda.vsnl.net.in

Abstract: India has been consistently making efforts to improve the quality of Agricultural Statistics more specifically Crop Statistics. But it has the unique model - **The Scheme for Improvement of Crop Statistics (ICS)** - for sustaining the quality. The model acts as a watchdog through well designed sample checks. The paper presents a brief picture of this model.

Key words: ICS, Sample checks on Area Enumeration, Area Aggregation and Crop Cutting Experiments, village.

1. Introduction

India has a federal structure with the state governments having primary responsibility of collection, compilation and publication of Crop Statistics. The Union government at the center, in its advisory capacity, is only concerned with the functions of coordination, consolidation and publication of statistics furnished by the state governments. The yield statistics are collected through the scientific technique of crop cutting experiments under the **General Crop Estimation Surveys** from as early as 1943-44. The area statistics are collected through complete enumeration of all the villages (lowest administrative units) in states where a primary agency exists at the village level and through sample surveys in the states where such primary agency does not exist. In certain other states, which do not fall under these two categories, these are collected through conventional methods. With the introduction of **Timely Reporting Scheme** in the former states and **Establishing an Agency for Reporting Agricultural Statistics** in the latter, the Union Government showed its interest in improving the timeliness of area statistics. Under these schemes, 20% of villages are enumerated completely on a priority basis for bringing out estimates of area in advance.

2. The scheme for Improvement of Crop Statistics (ICS) – The Model

ICS, the model, was introduced in 1973-74 with the main objective of locating the deficiencies in the state system of crop statistics and suggesting remedial measures to effect lasting improvements in the system. The model is unique in the sense that it completely devotes itself to the quality of data. Errors can creep into the crop statistics while **reporting** the crops and their area and **aggregating** them, as well as, during **conducting** crop cutting experiments. ICS takes care of these three sources of errors and guards against the adulteration in the quality through its three components of sample checks.

The stratified random **Sample Check on Area Enumeration** aims at studying deficiencies with respect to timeliness & accuracy of enumerating the crops and their area through physical verification of crop enumeration done by the primary worker in a sample of 10,000 villages in each recognized season. The check probes into qualitative aspects of area enumeration. It also makes quantitative assessment of the extent to which discrepancies are observed in crop & land use pattern and the resultant impact on area statistics. It brings out the possible errors in recording crop and crop-area viz, (i) Missing crops sown, (ii) Reporting crops not sown and (iii) Assessing area inaccurately.

The **Sample Check on Area Aggregation** embarks upon two types of errors - one at the time of aggregation and the other at the time of reporting such aggregated figures to higher authorities - in the very same 10,000 villages through verification of crop abstracts. Besides bringing out discrepancies at the village level, ICS also attempts sample checks on the aggregation at various administrative levels above the village level.

The stratified multistage random **Sample Check on Crop Cutting Experiments** covering a total of 30,000 experiments observes how far the primary worker adheres to the procedure prescribed in conducting the experiment. The check locates errors in selection of survey/sub-survey numbers, field, location of the experimental plot, measurement of the produce, etc. Unlike the checks on area, this is a concurrent check as mistakes, if any, committed by the primary worker in the whole process of the crop cutting experiments are corrected concurrently to the largest extent possible. It also specifies whether the mistakes observed were rectified or not. The check guards against the use of non-standard equipments in the conduct of the crop cutting experiments by specifying the status of both supply and use of crop cutting equipment. It helps to minimize errors occurring due to substitution of sampling units for one reason or the other.

3. Conclusion

The findings of ICS, which efficiently guards the quality of Agricultural Statistics, should be further fine-tuned to automatically refine the state estimates of area and production.

References

National Sample Survey Organisation, Government of India: Report of the status of estimation of crop production in India (1998-99)

National Statistical Commission, India: Interim Report (2001)

B.S.Minhas (1999): Indian System of Crop Statistics - Will the Cracks pull down the Edifice?

C.Rangarajan (2000): Restoring Trust in Statistics

6 June 2001

Contributed Paper Session

GIS AND OTHER TECHNOLOGICAL ASPECTS

Chair: R. Benedetti

Using GIS to Integrate Environmental, Social and Economic Issues to Assess the Environmental Impacts of Agricultural Policies

Isabelle Guinomet
GIM SA – Geographic Information Management
13 B, avenue Guillaume L-1651 Luxembourg
Tel +352 26 44 09 55, isabelle.guinomet@gim.lu

Keywords: GIS, impact analysis, visualisation, geo-spatial modelling

1 Introduction

New kinds of communication channels and information systems are needed for policy makers to address integrate environmental, social and economic issues to assess the environmental impacts of agricultural policies.

2 Why using GIS?

Impact assessment is a multi-disciplinary combination of physical and social sciences. That process identifies the future consequences of a current or proposed project, action, policy, plan or programme. GIS modelling provides tools and processing structures for incorporating spatial relationships which address "So What" in a decision-making context (map analysis and modelling). This extension of descriptive mapping to prescriptive mapping involves entirely new spatial reasoning concepts and procedures. The GIS techniques to be applied will include the use of data overlays, buffers and the creation of dedicated processing algorithms to create value-added information from the input data. Main use of modelling methods in a practical sense is :

- Synthesis and integration of data: Models can be used to generate missing data and overcome different levels of spatial scales and levels of aggregation.
- Forecasting models use information from the past to predict changes in the future. Such models can range from simple extrapolation methods to more sophisticated economic and mathematical models.
- Impact analysis: Common question of impact assessment is addressed by the 'what if' question. This approach is best suited for the requirements of the present proposal because it does permit to analyse the effects caused by a proposed or implemented policy decision.
- Visualisation: Geopolitical maps summarising the results of the modelling activities provide to decision-makers scenarios description with a synoptic and user-friendly

presentation. These geo-political maps are a good support to get a snapshot of future reality according to envisaged policies.

Using Information and Communication Technologies (ICT) as a catalyst, GIS gives the opportunity to join government, administration, researchers, public participation, education, and achieve connectivity between groups and network. This supports the need for transparency between project's opportunities and existing resources and should lead to raise synergy. Modelling to evaluate ex-ante and ex-post agricultural policies including a thorough assessment of their environmental impact Environmental address several key objectives - as stated in the table below:

Objectives	Key benefits provide by GIS (and linked motivations)
Develop a computer-based model or series of computer-based models for assessing the environmental impact of agricultural policies	GIS facilitate the running of the models (improvement of the understanding of agricultural policies impacts in the context of the global economy) especially at regional level
Develop the database required for using the model or models	GIS increase knowledge of environmental impact of agricultural policies (positive attitude or status is attached to policy makers)
Use the model to assess ex-ante and ex-post the impact of selected changes in agricultural policies on the environment	Geopolitical maps convey information about environmental impact of agricultural policies. Geopolitical maps increase the relevance of information to the audience (better policy making; sense of value at local, regional, national, EU level)
Develop a user-friendly interface	GIS interface facilitate awareness raising, transparency and better understanding between stakeholders; GIS interface allows the running of the models by politicians, administrators, and any other interested person having access to internet facilities.

2.1 Advantages of geo-spatial modelling

Spatial modelling includes the use of geographical (e.g. distribution of soil parameters in space) and non-spatial (i.e. statistical) data in a combined GIS / statistical analysis environment. Once the underlying principles of a phenomenon and its value as an indicator (e.g. the level of nitrate in the groundwater as an indicator for the intensity of agricultural practices) are described and validated, GIS techniques can be used to assess the phenomenon in its geographic distribution.

- To define a systemic context: Geo-spatial modelling allows taking into consideration site specificity, avoiding data aggregation at national level. It pays attention to the spatial dimension of problems encountered and allows regional differentiation and provides valuable insights for problems such as biodiversity, landscape, water quality and use.
- To consider the transboundary nature of environmental problems: A number of environmental problems between countries have impact of a transboundary nature.

Geo-spatial modelling have a potential to take into accounts social, economic and environmental nature, making evaluation of the trade-offs between the various objectives of a complex problem. Geo-spatial modelling support the definition of key indicators for policy makers to support agri-environmental decision-making.

3 Implementation of geo-spatial modelling

Geo-spatial modelling is carried out by taking the models and data into a Geographical Information System.

- Problem analysis: The use of different data sets and data sources in geo-spatial modelling raises a number of issues which need to be taken into account developing applications based on such data.

- Required data layers and potential data sources: In a geographic information system, information can be subdivided into two categories: base layers and thematic layers. Base layers are used for geo-referencing, spatial searches and background information. Thematic layers contain information on a specific thematic topic. The thematic layers are modelled according to the specific functional requirements for the use of these data. Additionally, geographic information is stored with a certain precision and spatial resolution. Depending on the purpose (e.g. display or analysis) a different resolution (cartographic scale) is required.
- Development of computer based models: The way policy measures affect the environment could best be described within frameworks such as suggested by Coleman. At macro level, policy makers promote measures. These measures will affect the environment within which the actor acts, on the macro level. At the micro level they will affect the possibilities of actors (farmers), resulting in a change in behaviour of the actors in a system. This change in behaviour again affects the environment at the macro level.

The Accuracy of Land Use Data in the European Union

Antonio M.D. Nucifora
DISEAE – Università di Catania, Italy. e-mail: antonio.nucifora@unict.it

Abstract: Data for the EU suggests that a remarkable increase in woodland has taken place in the post-war period. As clarified in this paper, however, this pattern is entirely determined by a problem with the French data. The abnormal increase in French woodland during the period 1965 to 1970 is associated with a change in survey methodology. It seems that the surface of woodland in France had been significantly underestimated prior to 1970.

Key words: Land use, Woodland, Methodological change, Inaccuracy.

1. Land use change in the European Union in the post-war period

Extreme caution needs to be exercised when using land use data for the European Union as a whole. Firstly, the data (in the Eurostat database) originate from different national surveys whose methodology has itself been changing over time. Secondly, the overwhelming dimension of France within the EU (about 35%) has potentially serious implications. Any statistical irregularities in the French data have the power to influence significantly the overall figures for the entire EU.

During the post-war period, more than 10% of the land in agricultural use in EU-9 has changed its designation. This has led to an increase of more than 15% in the surface covered by woods in Europe, and to an even faster increase in the areas destined for other uses. A preliminary examination suggests that the increase in woodland is concentrated wholly in the period between 1960 to 1970, while more recently the majority of the land released from agriculture has been absorbed by other uses.

2. A mystery of French statistics

The dis-aggregated data by country show clearly that this increase reflects almost entirely an increase in French woodland area. France contains almost 50% of the total EU-9 woodland and, as a result, the increase in the French statistics has a strong impact on the series for woodland in Europe. This increase of more than 3 million hectares (almost 30%) in French woodland in the 1960-1970 period needs further investigation. Some explanations have been suggested by Cavailhes and Normandin (1993), but these are in no way sufficient to explain such a dramatic change in woodland. The explanation proposed here is that the abnormal increase in French woodland during the period 1965 to 1970 is associated with a change in survey methodology. Surprisingly, Eurostat does not mention any problems with the woodland data. But it seems obvious that we are looking at an ‘artificial’ change in woodland in France. In fact, the methodology used to collect land use data in France changed in 1970. The French National Forestry Inventory (*Inventaire Forestier National*, IFN) was established in

1960 but did not produce full data for France until 1970. Before the introduction of the IFN, the assessment of the forest surface used to be derived from the properties registered in the land cadastre, on the basis of self declarations. The IFN results, on the other hand, are based on aerial photographs and may therefore be considered to be more accurate than data from cadastral surveys. From 1970 onwards, the data from the IFN have been regularly updated, with basically the same methodology. One plausible hypothesis, therefore, is that the official statistics were derived by extrapolating the intermediate years between two separate surveys for 1965 and 1970, thus showing a false gradual increase in woodland, and hiding the true spurious nature of the jump in 1970.

The reason for such a large divergence between the woodland areas spontaneously declared to the officials and those, appreciably larger, effectively encountered with the introduction of the appropriate surveying techniques, is probably the result of the fiscal dis-incentives existing against woodland in France. Every property in France is registered by the Cadastre Office and every land owner has to pay an annual tax which depends on the type of land. The French system taxes the land surface declared to the cadastral office with a rate which is higher for woodland than for any other land use; this is because there are no other subsequent taxes of wood production. In the light of this, the gap between the 1965 and 1970 figure for woodland in France might be mostly explained by the fact that some owners reported to the cadastre as agricultural land what was in fact forest.

This finding has been almost entirely neglected in previous studies on woodland in France and in the EU (Devaud, 1984; Lee, 1991; Bazire, 1993; Cavailhes and Normandin, 1993).[1] More generally, however, research studies on European land use change have not correctly identified the source of this abnormal increase in woodland, or have ignored it altogether. The result have been grossly incorrect analyses of past land use changes and significantly flawed forecasts.[2] This finding, therefore, reiterates the importance of a thorough examination of the data, particularly when dealing with land use statistics, which are notoriously subject to problems with changes in methodology.

References

Bazire P. (1993) La production des forêts françaises depuis 50 ans. *Forêt Entreprise*, 1993/94 (91), 6-10.

Best R.H. (1981) *Land use and living space*, first ed., Methuen & Co. Ltd, London.

Cavailhes J., Normandin D. (1993) Déprise agricole et boisement: état des lieux, enjeux et perspectives dans le cadre de la réforme de la PAC, *Revue Forestière Française*, 45 (4), 465-82.

Devaud F. (1984) Collections de statistique agricole, *Revue Forestière Française*, 36 (1), 11-21.

[1] An exception is Best: “Such a massive gain (in French woodland) over so short a period must raise doubts about the validity of the figure, and warrants further investigation.” (Best, 1981, pp. 177-78).

[2] See, for instance, the conclusions reached by Lee (1991, p. 14): “The EEC-9 woodland area increased by 3.9 million ha over the decade 1961-71. (…) If this rate of planting continued to 2000, an additional 8 million ha would be diverted to forest between 1985-2000.”

Lee J. (1991) Land resources, land use and projected land availability for alternative uses in the EC. In: Brouwer F.M., Thomas A.J., Chadwick M.J. (eds.) *Land use changes in Europe*, Kluwer Academic Publishers, Dorddrecht, The Netherlands.

Satellite Accounts: A Typological Approach

Franz Murbach
Swiss Federal Statistical Office, Agriculture & forestry section, Espace de l'Europe 10, CH-2010 Neuchâtel, Switzerland, e-mail: Franz.Murbach@bfs.admin.ch

Abstract: This contribution presents a typology as synopsis of a compilation study of different research work and methodological outcomes in the field of satellite accounts, mainly since the introduction of the System of National Accounts SNA 93. This attempt enabled to introduce into a coherent framework an initially heterogeneous enumeration of accounts, structuring the "System of Satellite and Economic Accounts for the Primary Sector", project launched by the Swiss Federal Statistical Office in 1997.

Keywords: Data Integration, Economic Accounts (primary sector)

1. Introduction

In 1997, the Swiss Federal Statistical Office (SFSO) launched a project named "SAKO-1, System of Satellite and Economic Accounts for the Primary Sector".
The aim of this paper is to present a typological approach which is being applied as a support for designing a comprehensive system of economic and satellite accounts for the primary sector (agriculture, forestry, fisheries). This typology is an attempt to summarise decades of methodological development in various countries and institutions, and has to be understood as a product of the analysis of the state of the art in this specific field.

2. Typological axes

We have chosen 3 main axes to specify a type of economic or satellite account, the reference (core) being the SNA 93 (respectively ESA 95):

Figure 1: *Typological axes*

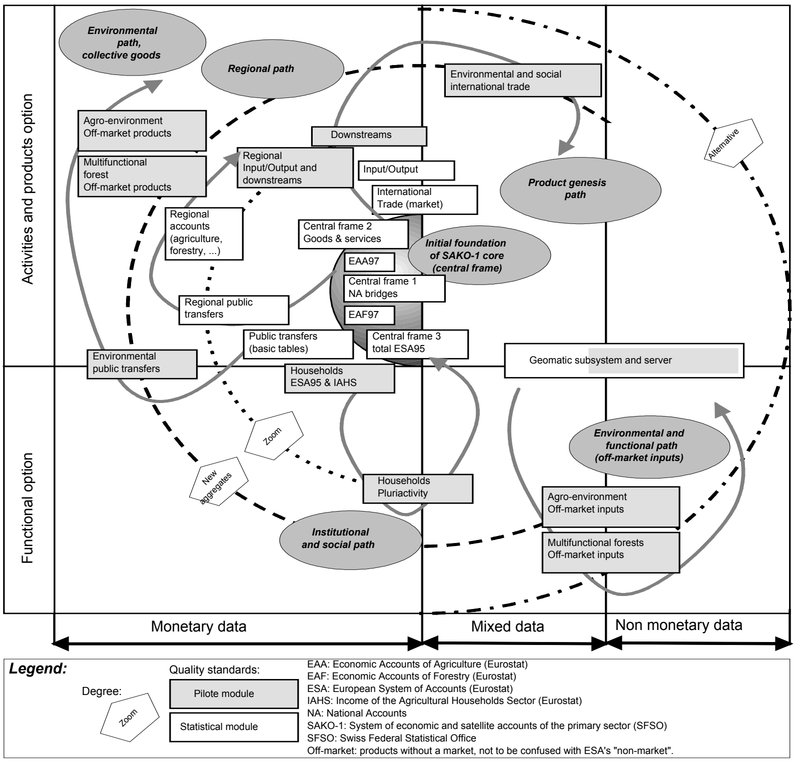
Degrees	Options	Zone
❑ Zooming	❑ Economic activities and products	❑ Monetary data
❑ New aggregates and classifications	❑ Functions	❑ Monetary and non monetary (mixed) data
❑ Alternative concepts		❑ Non monetary data

A further sense of depth and feasibility can be given to each account plotted:
- Pilot quality, not or only partly responding to the standards of public statistics;
- Statistical quality, responding to the standards of public statistics.

3. Plotting a design with the typology

The typological axis system is placed as underlay, on which the different economic and satellite accounts of the projected SAKO-1 system are plotted (application example). The core cannot be drawn completely, the truncated right part showing the "residual" (US NATIONAL RESEARCH COUNCIL 1999) left by the actual central frame concepts, where satellite accounts typically play their role. The different evolution paths bring the time aspect into the typology, and show links between the different modules.

Figure 2: *Application example: plotting the SAKO-1 system with the typology*



References

Office fédéral de la statistique, Projet SAKO-1, Concept global du cadre satellite (2000)*, (on request, only in french), containing a complete bibliography, (consulting by Mr. Gonzague Pillet, Ecosys SA Genève-Carouge),* Neuchâtel

US NATIONAL RESEARCH COUNCIL (1999), *Nature's Numbers – Expanding the National Economic Accounts to Include the Environment*. National Academy Press, Washington, D.C

Datawarehouse and Agricultural Statistics: Experiences

Nadia Rognon, Jean-François Fracheboud
Swiss Federal Statistical Office, Agricultural & forestry section, Espace de l'Europe 10, CH-2010 Neuchâtel, Switzerland
e-mail: Nadia.Rognon@bfs.admin.ch

Abstract: This contribution presents an experience of implementing a data warehouse gathering agricultural statistical data in order to insure their dissemination. The data warehouse is outlined through its structure, data model and metadata backbone. The issue dealing with the interactions between micro-, macro- and metadata is discussed. The possibilities of data dissemination and availability are presented.

Keywords: Data Warehouse, Metadata, Dissemination of data

1. Introduction

The aim of this paper is to present an experience of gathering agricultural statistical data into a data warehouse in order to insure their dissemination. It presents a project which is presently realised in the Agricultural and forestry section of the Swiss Federal Statistical Office and which aim is to gather all statistical data about agriculture and forestry in a common container in order to facilitate their dissemination and to guarantee their permanence.

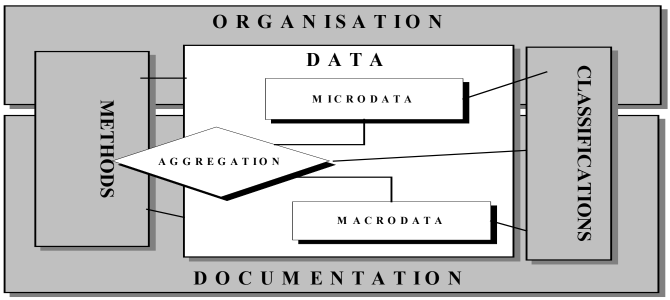
2. Data Warehouse structure and data modelling

The data warehouse is structured into several parts that interact as presented in the figure 1 below. The data themselves are stored in a quite simple, but universal data model. In this data container are stored both microdata and macrodata. The term "microdata" is used for the finest individual data, that can be the final data of a census for example. These data are clean and definitive and they will never be updated. The "macrodata" are the result of a calculation that is worth to store. These data are calculated once and they never change, like the microdata. The data model of micro- and macrodata is structured in a way to optimize the querying and the aggregation of data. In the practical example presented here, the data container is a relational database and the RDBMS (Relational Database Management System) used is Oracle™.

3. Metadata and their interactions with micro- and macrodata

Metadata is very important to be able to make the best use of the data warehouse. It covers a wide range of information about the data that are stored, their quality, the sources of data, which data are available, the classifications used to code the data, etc. In Switzerland, the amount of metadata in agricultural statistics is quite important, as there are about 300 statistical sources and 70 data producers with no achieved harmonisation between them. In figure 1 below, metadata is represented in grey. It provides several ways of helping the user to know what to find where, which is of primary importance because the amount of data stored can be huge. In the case presented, metadata is presently stored in an Oracle™ database, but a specific application for managing some of the metadata (classifications, organisation) is developed in co-operation with other statistical offices and will be available soon.

Figure 1: *Structure of the data warehouse*



4. Dissemination of data

As the data are stored at the same place and are regarded as a universal model, their dissemination is very flexible. The data can be accessed by commercial data mining software (Business Objects™, SuperSTAR™, …), or by direct queries on the database. The dissemination via internet is also an interesting solution that will be developed in the future.

5. Conclusion

The use of a data warehouse in statistics is undoubtedly very valuable to assure the permanence of the data. It is also a great help in data dissemination and provides a wide range of possibilities in this domain.

References

Inmon W. H. (1996) *Building the Data Warehouse, Second Edition*, Wiley Computer Publishing, New York.

Statistics Canada (1999) *The integrated metadata base of statistics Canada*, contributed paper in Conference of European Statisticians, UN/ECE Work session on Statistical Metadata, Geneva, Switzerland, 22-24 sept. 1999.

An Agricultural and Environmental Information System in Intensive Agricultural Areas[1]

Luis Ambrosio
Luis Iglesias
Carmen Marín
Universidad Politécnica de Madrid. Departamento de Economía y Ciencias Sociales Agrarias. E-mail: flores@eco.etsia.upm.es

Julio Montañés
Luis Alberto Rubio
Junta de Andalucía. Consejería de Agricultura y Pesca. Servicio de Estudios y Estadísticas. E-mail: lrubio@cap.junta-andalucia.es.

Abstract: This communication includes the main points of an information system on the agrarian complex and its environmental impact, designed for the intensive crop (horticultural) areas in Almería (Spain). The basis of the system developed is an area frame sampling integrated in a geographical information system (GIS).

Keywords: area frame agricultural surveys, residuals, environmental impact, GIS.

1. Introduction

In specific areas of intensive agriculture the amount of residuals obtained is a factor limiting production development. In these areas, for the design and follow-up of agricultural policy measures, it is especially important not only to have the traditional information about production, yield and production factors, but also information about the environmental impact of the agricultural activity. On the other hand, the agricultural activity in these areas is very much integrated in the food and agricultural complex along with agricultural input supply companies and those of the transformation and marketing of agricultural products.

2. The system bases

In addition to the area frame, with physical boundary segments [FAO (1996)], as a basis for the selection of a segment sample and of a sample of farmers, the GIS integrates a list frame sampling for the selection of a sample of marketing agents and another for the selection of a sample of recycling companies. The programmes for the estimation calculation based on the information extracted from the samples are also integrated in the GIS.

[1] Finantial support from Junta de Andalucía (Spain) for Project P98022095 is gratefully acknowleged

3. The questionnaires

Together with questions related to (i) crop alternatives, (ii) production, yield and the calendar of crops as well as (iii) production factors, especially water, the questionnaire for the survey of farmers includes others concerned with (iv) product marketing and (v) wastes generated (plastics, stubbles and damaged fruit unsuitable for marketing).
The questionnaire for the survey of marketing agents contains questions about the distribution chain, logistics, the type of contractual relationship and product standardizing (traceability). The questionnaire addressed to recycling companies focuses on recycled amounts and products obtained from wastes.

4. Estimations and analysis

In the survey of farmers, the dichotomous variables associated with some of the questions were treated as continuously associating the area of the surveyed farmer's agricultural exploitation with value 1 of the dichotomous variable and otherwise zero: in this way the estimation of proportions are in fact estimations of the proportion of the total surface cultivated by the farmers whose answer is associated with value 1 of the dichotomous variable. This transformation is necessary as the frame used for the selection of the sample is only that of the area, since the list frame on the group of farmers with a minimum coverage is not available.
The precision of the estimations related to marketing and environmental aspects does not differ significantly from those common to the area frame sampling methods for the estimation of characteristics related to agricultural production. We therefore believe that extending the traditional estimation procedures to these other food and agricultural as well as environmental aspects are not especially difficult.
The survey of marketing agents and recycling companies is only based on a list frame and is not accompanied by additional difficulties.
Based on the GIS some spatial variables used in the data analysis (distances of the plots to the irrigation channel, to the selling points of the products and to urban centres where municipal dumps are located) were determined.

References

FAO. (1996). *Multiple Frame Agricultural Surveys. Vol. 1 and 2*. FAO Statistical Development Series. Nº 7. Rome.

Reasons for an Agricultural Service Web-Site

Aurora De Santis
Italian National Institute of Statistics, Via Ravà 150 – 00142 Roma, Italy
e-mail; aurora@istat.it

Abstract:According to the need of improving quality of statistical information, the Agricultural Service started a complete revision of contents, surveys and processes within an integrated system that will include, in the near future, a specific web-site with the declared aim of improving quality in the timeliness and accessibility of data-dissemination.

Keywords: official statistics quality, information technology opportunities.

1. Reasons for a statistical web-site

Up to some years ago, quality concepts in statistics were mainly based on accuracy. Nowadays there is a considerable enlargement as regards this classical vision: not only accuracy but also transparency of the process, together with completeness, comparability, coherence, relevance, accessibility and finally punctuality in disseminating results play a role as well (Eurostat 2000). The balance between these components is entrusted to statisticians who should know who are the final users and what are their needs and priorities. Thanks to internet-culture, these priorities are changing, and statistical data is more and more important in understanding the various phenomena of society, in public decisions and policy. Accordingly, the aim of official statisticians to enable a large number of people to access and share official data becomes relevant and the "the net" has the particular merit of presenting statistical information dynamically and making it understandable and available to everyone at very low cost. In addition to that, monitoring the number of access to the web-site, a statistician can also have a feedback about the interest on his activity and the preferences of users. In this way, a statistical web-site could become the prime mode of delivering official statistics and of managing statistical enquiry.

1.1. Main advantages

The statistical process and data access benefits, obtained from a fully integrated web-site architecture, will be: extending public documentation at low cost, reducing communication costs and cycle time for new scheduled activities, linking with any relevant organization, improving the process and monitoring performance. The main advantage is that data could be available without any kind of mediations: Internet becomes the most important source and, at the same time, instrument of research.

1.2. Main implication

The statistical web site would be fully integrated with statistical systems and processes so that the costs of transferring information to the web-site are very low. This integration includes: technology, tools, systems, processes, reporting and so on. This is what the Agricultural Service has done over the last years: it has achieved a complete revision of the Italian agricultural statistical system towards new strategic lines regarding methodological, managerial and technological areas (Martino, De Santis & Salvi 2000).

2 How it could be organized

A typical web-site might consist of five quite distinct elements: picture about agriculture, information gathering, secure access to confidential unit record, documentation, other services (e.g. a communication channel and so on). What kind of site one can create depends on "the philosophy" you have about it: basically you can have a mixture of these five points in three distinct architectures. The first one will be focused on immediately usable material, free access to the entire documentation and a user-friendly structure of the web-pages by prefab HTML files. This kind of web architecture is quite easy to implement but it is not able to support a user query any better than time series data do, because of its rigidity. All statistical web-sites have this scheme (e.g. ISTAT and NASS). The second one is based on a database of final estimates and involves a figure with the task of collecting data which allow for a different way of working (e.g. FAO). The third hypothesis is based on a data warehouse that supports the informative process providing a platform of integrated data whose elements can be organized according to many different requirements. Users can partly customize both structure and contents, selecting only subsets of data of their interest. Whatever the choice may be, the Agricultural Service web-site has to contain metadata of our surveys and the data of the most important surveys which could improve the comprehension of our results (e.g. data on prices of agricultural product, on farms, on labour, on foreign trade), which is more than just providing representative links of national and international sites (e.g. EUROSTAT, FAO, USDA/NASS, OECD, UN-ECE, MIPA, INEA, ISMEA, etc.).

3. Final remarks

Internet Technologies give statistical offices the chance to be a forefront of the public sector modernization. ISTAT is deeply involved in this challenge: the aim of the National Statistical System (SISTAN) is "promoting every action aimed at spreading culture of technology and disseminating official statistical data, for everyone". The Agricultural Service is taking part in this change, introducing, in the revised system of agricultural statistics, the issue of its own web-site, which now in an experimental version is shown in http://intranet/~agr/index1.htm.

References

Eurostat (2000) *Assessment of quality in statistics. Item 4 of the agenda "Standard quality report"*. Luxemburg, 4-5 Aprile 2000.

Martino L., De Santis A., Salvi S. (2001) *An integrated system to improve quality in official statistics*. Proceeding of International Conference on Quality in Official Statistics. Stockholm, 14-15 May 2001.

Coupling Remote Sensing Techniques with GIS Tools for Change Detection Analysis of Landscapes Related to Human Disturbance

Giovanni Zurlini
Dipartimento di Biologia, Laboratorio di Ecologia del Paesaggio, Università degli Studi di Lecce, Italy, e-mail: vico_vico1@excite.com

Alessandro Ferrarini, Pierfrancesca Rossi, Nicola Zaccarelli
Unità di Ricerca CNR, Progetto "Conoscenza, Conservazione e Gestione della Biodiversità", Italy

Lamberto Soliani
Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Italy

The integration of socio-economic and environmental data from different sources into GIS (Geographic Information Systems) nowadays represents the main strategy for the analysis of human population dynamics and the assessment of their impacts upon ecological systems in order to provide different alternatives for their sustainability. The association between statistical and geographical data plays a fundamental role to build standard methodologies to be shared by the European statistical system, and it is the basis for converting a raw data collection into an organized structure to allow an appropriate use of available information archives. Data are usually planned to be aggregated to represent municipalities or larger administrative areas starting from micro-territorial analyses. In urban areas, the analysis of territory requires a deeper and more detailed knowledge, so that every building can be identified. For residential areas, new census units, every single family, together with their components as well as regularly and occasionally residing people can be located at the moment of the census. Transfer from a highly detailed to a broader scale resolution cartography has now become relatively easy and rapid with great advantages to scientific research which surely requires georeferenced and logically integrated information organization at different scales.
Statistical analyses of census data in 1991 were pertaining to a map of Italy mainly derived from satellite remotely-sensed images of the whole national territory, but with a resolution unsuitable for applications to large urban areas. This limitation was determined by satellite sensors available at that time, which provided a comprehensive but rather coarse resolution, in the total absence of digital aerophotogrammetric cartography either in raster or vector format. Nowadays it is possible to answer to the several needs resulting from landscape management, environmental impacts, soil defence, civil protection and official statistical analyses which require digital geographical supports at relatively higher spatial and temporal resolutions. As to recent censuses, the use of digital ortophotos coupled with GIS makes possible the acquisition, processing, analysis, storage and graphically and alphanumerically rendering of a variety of territorial data. To this purpose, georeferenced objects linked to attributes, geographical, alphanumerical and statistical tables can be used. Spatially explicit

functions, such as e.g. distance and areal measures, centroid search and point to line least distance are easily manageable.
Multi- and hyperspectral remotely-sensed data act today as a fundamental information source for the landscape classification inside GIS (Wilkinson, 1993). Several applications in this area are well documented in the scientific literature.
The national Project "Carta della Natura" (Rossi O., Zurlini G., 1993; Zurlini et al., 1999), promoted by the Italian Ministry of the Environment and involving a variety of researchers from different italian universities under the coordination of the University of Parma, aims at generating digital thematic maps as Geographic Information System (GIS), coverage in polygonal (vector) format of (1) existing land cover habitat types and biodiversity component distributions, and (2) different kinds of man-induced disturbance.
The project, akin to the Gap Analysis Program in the States, aims at identifying gaps in the existing reserve network in order to establish new reserves and protected areas to get a more representative network of regional biological diversity, based on (1) their "natural values", and (2) "fragility". This project makes use of traditional ground data as well as multispectral and hyperspectral data such as satellite and Mivis images, and colour digital ortophotos.
The basic rationale of the Project is that conservation and management of environmental resources mainly depend on the successful gathering, digitalization, storage and integration of data and knowledge coming from a variety of sources and on their subsequent processing and cartographic rendering. As an illustrative example, results from a study carried out within the framework of the Project "Carta della Natura" on the Baganza stream watershed with remote sensing techniques (Ferrarini et al., 2000) are reported. Change detection analysis as well as spatial analysis of the landscape are the main issues of these investigations. Image algebra is an approach to change detection where the intensity and direction of changes are quantified by differencing or ratioing two remotely-sensed images acquired at different times (Fung et al., 1988; Green et al., 1994). Those change detection techniques were applied to NDVI index variation over 5 year period from Landsat TM scenes of the Baganza stream watershed. NDVI is one of the most used remotely-sensed vegetation indexes and is calculated as the normalized difference ratio between red and infrared spectral bands. NDVI index has been extensively used to measure vegetation biomass and vigour on a wordlwide basis. Foreign countries routinely use averaged NDVI information to assess crop production and health over wide regions and provide crop forecasts. It is indeed strongly correlated to health conditions of vegetation cover like nutrients availability and soil water availability. Interpretation of NDVI change detections was supported by the mosaic of CORINE habitats. In fact, 2,334 relatively homogeneous patches were recognized as CORINE habitats (Rossi P., 1999) in the stream watershed according to the criteria stated by the European Community, which aim at compiling an inventory of habitat of major importance for nature conservation in the European Union to provide a unique evaluation basis for conservation evaluation and management. Landscape mosaics of CORINE habitats were identified on the basis of vegetation covers, soil types and land forms and coded at different levels of a hierarchical system of mosaics of patches, within patches comprising very broad sintaxa at the landscape level down to alliances and associations. By comparing detected changes in relation to habitat mosaics, a) changes due to human disturbance in the lowland, b) abandonment of human activities in the mountain, c) natural phenomena dynamics on short temporal terms, were

identified and spatially appreciated. Results stress the need to couple remote sensing techniques with GIS tools in order to generate change detection maps which can give policy makers and land managers information to foresee the effect their land use decisions will have on existing processes and risks to biological conservation.

References

Ferrarini A., Zaccarelli N., Rossi P., Cristiano D., Orlandini L., 2000:"*Change detection* degli Habitat CORINE nel bacino del Torrente Baganza (Parma) attraverso l'indice NDVI", *Rivista Italiana di TELERILEVAMENTO, 19*.

Fung T., LeDrew E. (1988): The determination of optimal threshold levels for change detecion using various accuracy indices, *Photogrammetric Engineering & Remote Sensing*, 54 (10): 1449-1454.

Green, K., Kempka D., Lackey L. (1994): Using remote sensing to detect and monitor land-cover and land-use change, *Photogrammetric Engineering & Remote Sensing*, 60 (3): 331-337.

Rossi O., Zurlini G. (1993): Primi elementi conoscitivi essenziali per la realizzazione della Carta della Natura (Legge n. 394 del 6/12/1991), *S.It.E. Notizie, Bullettin of the Italian Ecological Society*, Vol. XIV, 1993: 46-56.

Rossi P. (1999): Analisi della diversità vegetazionale in Val Baganza mediante la classificazione CORINE Biotopes, Tesi di Laurea in Scienze Biologiche, Facoltà di Scienze Matematiche Fisiche e Naturali, Università di Parma.

Wilkinson G., 1993. *Neural Networks: a new analysis tool for GIS and remote sensing*, GIS Europe, 2, 38-40.

Zurlini G., Amadio V., Rossi O., 1999: "A Landscape Approach to Biodiversity and Biological Integrity Planning: the Map of the Italian Nature", *Ecosystem Health*, 5(4), pp. 294-311.

7 June 2001

Contributed Paper Session

SOCIO-ECONOMIC ISSUES

Chair: M. Steiner

The Effect of Aging Farm Holders on Gross Income and Agricultural Activity in Israel 1971-1995

Elana Dror
Central Bureau of Statistics – Israel, Elana@cbs.gov.il

1. Background

The "Moshav", a semi-collective rural locality consisting of family farms began to develop in Israel in 1921, the year the two first Moshavs were established. Over the years there have been changes not only in the agricultural development in general but also in family farms on the Moshavs. Some of the changes occurred as a consequence of changes in the economy of the state of Israel in the process of its development such as growing wages and family income growth. Some of the changes occurred in the family farms in the agricultural aspect, such as: vast improvements in agricultural technology (special machines that help plant and pick the crops), and intensity of the farm holder's work on the farm (partial work on the farm and more work off the farm).
Many of these changes are still occurring in the rural society. The number of active farms is decreasing, yet the population living in the rural society is increasing. The main reasons for that are the needs and interests in improving housing environment away from the city and its aspects.
Today there are 400 Moshavs in Israel and the main question is how many Moshavs will be able to adjust to the changes and still exist in the new format in the near and far future.
The following is an effort to explore the changes in the family farms in Moshavs in the past twenty-five years. Knowing about these changes might help predict what might happen in the family farms of the Moshavs in Israel in the future.

2. Methodology

2.1 Number of farms

Within the framework of this research project, data was collected from the two agricultural censuses, in 1971 and 1981, and from the 1995 agricultural survey. After merging data on farms in Moshavs from all three periods the file contained 685 farms, surveyed in all three surveys. The farm ID was changed in certain farms and therefore this parameter could not be used to merge the files. Parameters that appeared in all three files were used to merge them. The parameters, "year of birth" of the farm holder and the "year the farm was received" (or founded or bought) were chosen. The final study file contained 685 farms in which the farm holder has not changed for 25 years.
The farms in the file represent active and non-active farms. An active farm is a farm that has a Gross Value Added (GVA) of over $1000.

2.2 Gross Income Per Day compared over the years

GIPD (Gross Income Per Day) was estimated as follows: GVA of the year 1995 was divided by the number of workdays on the farm in 1995. The result was divided by the GVA for 1981 divided by the number of workdays on the farm in 1981.

$$GIPD = \left(\left[\left(GVA\ 95 /(work\ 95)\right)/\left(GVA\ 81 / work\ 81)\right)\right]\right)$$

The final estimate GIPDPY (Gross Income Per Day Per Year) is the growth percent per year:

$$GIPDPY = \left(\left[\sqrt[13]{GIPD}\right] - 1\right) * 100$$

The same estimates were calculated also for the farm holder, but instead of taking all the workdays in the farm only the workdays of the farm holder were counted.

3. Results & Conclusions

The farm holders were divided into three main groups according to their age:

Table 1. *Number of Farm holder's and GIPDPY according to age in 1995*

	Number of farms	GIPDPY(Whole farm)	GIPDPY(Farm holder)
Total	685 (100%)	8.5%	2.0%
Farm holder's age is over 70	96 (14%)	8.3%	0.75%
Farm holder's age between 70-46	492 (72%)	8.1%	1.29%
Farm holder's age is younger then 45	97 (14%)	11.0%	4.22%

The table depicts the following conclusions:

1. The GIPDPY for the farm holder is correlated highly with his age. The younger the farm holder the higher is his GIPDPY.
2. We see that the GIPDPY for both groups are the highest in the youngest age group. The obvious reason is that the farm holder in this age group works more days on the farm then the farm holder in the older age groups. We might have thought that the younger farm holder would work more days off the farm (outside the farm) and this might have caused the GIPDPY to be lower in this group but that is not the case.

The aim was to examine how changes in the farm itself and the aging of the farm holder affected the indexes. The main conclusion is that farms with older farm holders do not do as well as farms with younger holders. This conclusion is not obvious, since we could have thought that older farm holders have a more established farm and therefore have a better gross added value than farms of young farm holders.

The current database was too small to enable in depth investigation on how changes of farm holders affect the indexes. Currently families that live on a farm do not only make a living from agriculture but have external income sources as well. Therefore further studies are important in order to examine the farms' income from work outside of the farm and from welfare support.

Empirical Analysis of Profitability of Agricultural Enterprises in Kyrgyzstan

Roman Mogilevsky
Center for Social and Economic Research in Kyrgyzstan (CASE-Kyrgyzstan)
P.O.B. No.696, 720017, Bishkek, Kyrgyz Republic
Tel: +996-312-471510, e-mail: rmogilevsky@hotmail.com

Abstract: The paper contains results of analysis of profitability of agricultural enterprises in Kyrgyzstan based on survey data. The main hypothesis under consideration is that there is a systematic divergence in profitability between different types of enterprises in transitional agriculture of Kyrgyzstan. Empirical analysis has shown however that differences in profitability are statistically insignificant. The survey data gave an insight into some common problems of Kyrgyz farms irrelative of their type impeding development of efficient enterprises.

Keywords: Economic Transition in Agriculture, Efficiency of Agricultural Enterprises

Agriculture is the most important sector of Kyrgyzstan's economy. Annually it produces about 40% of the total GDP of the country (38% in 1999). Agriculture employs 52% of all employed people, 65% of the population of the country lives in rural areas. During the recent years, Kyrgyzstan is actively reforming its agricultural sector. These reforms include introduction of private ownership for land, creation of private agricultural enterprises, transformation of organizational and legal forms of agricultural enterprises, ways to provide agriculture with the necessary production resources and equipment, sale of the produce and pricing of products. It appeared that a large number of various types of agricultural enterprises differ in size, type of ownership on land, livestock, equipment and other resources, availability of these resources, as well as in management forms and principles of distribution of results. All enterprises can be divided into four main types: (i) individual farms (small farms formed by only one household, with a typical size of a land parcel from 1 to 10 hectares); (ii) peasant farms (farms created by several households, usually blood-related, with a size ranging from 5 to 50 hectares); (iii) collective farms (large farms, successors of former kolkhozes, their typical size is from 100 to 5000 hectares); and (iv) state farms (specialized farms – seed-growing, experimental, service farms, their size is similar to that of collective farms).

The main aim of this research was to collect and analyze the information on profitability determinants for agricultural enterprises, and, in particular, on the influence of the organizational and legal form of the farm on its effectiveness indicators.

This work is based on the results of a sample-based survey of agricultural enterprises in the Kyrgyz Republic. During the period of November 1999 – February 2000, 468 farms were surveyed that represent all types of property; they comprises about 1% of the total number of agricultural enterprises in Kyrgyzstan. The survey was conducted through interviews based on a questionnaire designed specifically for the this purpose, the questions of which covered the main aspects of economic activity of agricultural enterprises in 1999 calendar year.

This research is based on comparison of farms of various types by the profitability level, which is considered to be a generalizing indicator of farms' efficiency. A comparison of profitability indicators shows that there is no statistically significant difference in the profitability of different types of the farms. The net margin indicators per hectare of land and per worker have approximately the same value for all types of farms. The existing differences in profitability are related more to the regional differences (more associated with climate and the respective agricultural specialization) than to differences in the types of farms. It is obvious that if the organizational and legal form of a farm does not significantly influence its profitability, and that there are other factors determining the size of profit, for example, the land quality, specialization of the farm, etc.

An analysis of farms costs shows that their structure in many ways explains low farm profitability rates registered by the survey. The biggest share goes to the costs without which a farm just cannot exist: seeds, forage, and mandatory payments. The share of costs ensuring an intensive growth of production (chemical protection of plants, improvement of livestock breeds, veterinary needs, etc.) is not large, and often quite insignificant. The farmers participate very little in supporting the production and social infrastructure necessary for them, and this means that long-term development of this infrastructure cannot be secured. The production resources markets almost do not function. It is obvious that any increase of farm profitability is hardly possible without resolution of these priority problems.

Regressing of farm profitability on a set of explanatory variables allowed revealing that the most influential factors of effectiveness of agricultural enterprises are number of workers per hectare, share of irrigated land in farm's total land area, and number of milking cows.

The conducted statistical analysis allows us to make the following conclusions about the factors determining profitability of surveyed farms in 1999 in Kyrgyzstan:

- a large role is played by the structure of crops and livestock: selection of the most profitable cultures allows to increase the profits of farms significantly;
- in many cases the farms have an opportunity to increase profitability as a result of switching to more labor-intensive types of production;
- the quantity of irrigated arable lands and irrigation problems in general have a decisive significance for profitability of agricultural production;
- under the conditions of underdeveloped markets, provision of farmers with credits, even though it has a certain significance for growth of profitability, does not play the role that could be expected.

A general conclusion that can be drawn from everything said above is that the market mechanisms in agriculture are far from working efficiently and at full capacity, and therefore their development is one of the most urgent and high priority tasks.

Difficulties Encountered by Farm Workers' Families in Accessing some Social and Health Services in Italy

Mario Bolzan, Maria Marchesan, Silio Rigatti Luchini
Dipartimento di Scienze Statistiche, Università di Padova
Via Cesare Battisti 241, 35121 Padova, Italy
e-mail:mbolzan@stat.unipd.it

Abstract: We report some results of a new fuzzy approach to measure the difficulties in accessing some important health and social services by farm workers' families. The work is based on a function that takes into account both the level of difficulty of the single family and the observed frequency of that level in the population. The analysis is carried out on data-base of ISTAT Multipurpose Survey of 1997 in Italy.

Key-words: Fuzzy sets, Farm Workers' Families, Access to Health and Social Services.

1. Introduction

Exploiting the theory of fuzzy sets (Kruse et al., 1994) allows us to construct functions expressing the membership level of a single family group who experience some kind of hardship (Giles, 1988). A recently presented fuzzy measure (Bolzan, 2000; Bolzan & Rigatti Luchini, 2000) can measure the degree of difficulty encountered in acceding to some social and health services in Italy and large territories by farm workers' families, i.e., the agricultural sector (ISTAT, 1998).
The data-base is composed of more than 10,000 families of the ISTAT multi-purpose survey of 1997, covering aspects of daily life in which, among other things, the degree of difficulty (three levels: none, slight, great) encountered by families who resorted to eleven types of services offered in the territory in question. The measures of difficulty expressed by fuzzy function Dij vary between 0 and 1, indicating degrees of membership of the family in gradually and increasingly difficult conditions.
Dij is the normalized distance separating the condition of the i-th family (jointly expressed by the distribution function of the variable difficulty encountered in accessing the j-th service, and by the distance of the level of the same variable with respect to the lower level) and by the situation of minimum difficulty measured in the population with regard to the j-th service. It therefore expresses a distancing from the 'best' situation, i.e., a measure of observed difficulty.

2. Some initial results and reflections

Table 1 lists the values of fuzzy function Dj, the mean of measures in all families regarding service j, subdivided by geographic area and type of service. Clearly, the 'agricultural' family expresses a set of difficulties which are definitely greater than those reported by 'other' families, whose family heads are not occupied in agriculture (sometimes even 50% more than the others; see, in order: Pharmacy, Creche, Kindergarten, Primary School, Secondary School, Food Shops and Supermarkets). This condition also emerges, although to a more intense extent, in various areas in Italy where differences unfavourable to agricultural families are recorded - in the North-West, see, for example, Food Shops and Supermarkets. The reasons for these greater 'difficulties' are ascribed to the fact that, by their very nature, these services are preferentially concentrated in geographically limited areas, in urban and industrial centres. In southern Italy and the Islands, farm workers' families record alternating tendencies: in the islands, values are lower than the national mean and, vice versa, higher in the South. However, these differences may also be due to different sensitivity in perceiving conditions of difficulty.

Table 1: *Mean values of degree (Dj) of difficulty in accessing some services in Italy for farm workers' families (upper value) and other families (lower value).*

TYPE OF SERVICES	NORTH/ OVEST	NORTH/ EAST	CENTER	SOUTH	ISLANDS	ITALY
PHARMACY	.225	.191	.216	.243	.172	.213
	.130	.109	.134	.160	.146	.136
FIRST AID	.329	.335	.314	.446	.399	.377
	.312	.291	.293	.367	.316	.317
POST OFFICE	.165	.153	.197	.239	.197	.198
	.137	.103	.156	.200	.183	.154
POLICE	.292	.250	.200	.314	.224	.266
	.227	.190	.204	.263	.188	.220
TOWN HALL SERVICES	.163	.188	.213	.270	.179	.216
	.176	.149	.217	.223	.208	.193
CRECHE(*)	.298	.220	.277	.128	.174	.177
	.129	.108	.098	.096	.138	.109
KINDERGARTEN(*)	.219	.147	.265	.123	.112	.141
	.084	.085	.092	.100	.118	.094
PRIMARY SCHOOL(*)	.138	.197	.221	.149	.105	.147
	.086	.103	.089	.109	.107	.099
SECONDARY SCHOOL(*)	.300	.202	.294	.202	.152	.205
	.154	.138	.134	.146	.140	.143
FOOD SHOPS and MARKETS	.227	.189	.171	.206	.139	.182
	.137	.100	.116	.135	.111	.122
SUPERMARKETS	.320	.270	.260	.310	.180	.270
	.208	.175	.177	.213	.140	.190
No. families	99	183	101	319	138	840
	2449	2257	1895	2870	1020	10491

(*)Question only posed to families with children attending such schools.

References

Bolzan M. (ed.) (2000), Nuove misure del disagio familiare e di accesso ai servizi socio-sanitari per condizione di salute, *Seminar Proceedings*, Padova, 6 April 2000.

Bolzan M., Rigatti Luchini S. (2000), Morbosità acuta e stato di salute percepito: primi risultati di un approccio fuzzy per misurare la salute, 40ª Riunione Scientifica della Soc. Ital. di Statistica, Firenze, 26-28 April 2000, *Proceedings*: 663-666.

Giles R. (1988), The concept of grade of membership, *Fuzzy Sets and Systems*, 25: 297–323.

Kruse R., Gebhardt J., Klawonn F. (1994), *Foundation of Fuzzy Systems*, John Wiley & Sons, England.

Socio-Professional Self-Regeneration of Active Agricultural Population of Serbia by Data of Yugoslav Population Census, 1991

Vladimir Stankovic
Republic Statistical Office of Serbia

Mileva Zizic
Faculty of Economy, Belgrade, Yugoslavia
11000 Belgrade, Milana Rakica 5, tel. +381 11 411 836, fax 381 11 411 260
E-mail: filippi@szs.sv.gov.yu

Abstract: In the three last decades, between the pilot 1960 census and the last Yugoslav population census 1991, a number of big socio-economic changes occurred, and one of these is the phenomenon of rapid weakening of inter-generation socio-professional self-regeneration of active agricultural population which is in close correlation with dynamic processes of industrialization and urbanization.

Key words: census, social mobility, active agricultural population

1. Introductory notes

The Yugoslav census of 1991, among other things, is characterised by its authentic material that contains some basic information on inter-generation socio-professional population mobility. However, the first attempt in that field in former statistical practice was undertaken in 1960, when, within the expanded program of pilot population census contents, collection of information on social mobility in Yugoslavia on the sub-sample of 80 statistical districts was carried out.

In the 1991 census, questions on father's and mother's occupation were asked to the total censused population, regardless the age, activity, literacy and other individual characteristics.

Although the official statistical classification of activities, which first of all respects functional contents of the professional activity, is not suitable for more subtle sociological investigation of vertical socio-professional mobility, yet within the frame of this classification scheme, the basic rising and falling directions of inter-generation socio-professional movements of censused active population could be noticed quite clearly.

2. Level of socio-professional self-regeneration of active agricultural population

In earlierr investigations of social mobility agricultural layer it became clear, on one side, that there are rising social movements on the pyramidal scale of social position, and on the other side, that there were no significant downward trends, so that the scope and social survival of this layer nearly exclusively depends on the intensity of socio-professional self-regeneration, which slackened as the time was passing.

Along with self-regeneration dwindling due to a "run away" from agriculture, which, due to the results of post-war censuses was continuously decreasing, this had a crucial influence on farmers' descendants in the formation of an agricultural layer. So, as early as the1960 pilot census, it was stated that over 9/10 (92,6%) of farmers were then of agricultural origin. If we disregard area and methodological differences with respect to the scope of observed population, the 1991 census gave essentially no different picture. In the current composition of the agricultural population in Central Serbia, farmers' descendants make up about 9/10 (88,4%) of the population, while in the identical composition in Vojvodina their presence is over 8/10 (83,2%), while data for Kosovo and Metohia are not available due to the 1991 population census boycott by Albanian population.

As regards the intensity of the active agricultural population self-regeneration, quite significant changes could be noticed in this thirty years period. Self-regeneration was, in the pilot 1960 census much more outspoken (63,8%) than the 1991 census results show for Central Serbia (50,2%) and for Vojvodina (36,5%). Thus, the significant majority of censused persons (about 2/3 of agricultural social origin) in 1960 remained in the layer where they belonged by birth, but at the beginning of 1990's there is complete numerical balance between those who have "inherited" agricultural occupations from their fathers and those who have these occupations left. In that respect, the situation in Vojvodina is even more specific: three decades ago only 36,2% of active with agricultural origin did not participate in self-regeneration of its original social layer, while in 1991, approximately the same number of active (36,5%) remained in socio-professional layer of their fathers.

At the beginning of 1960's among non-farmers there were more numerous persons with an agricultural origin (51,0%) than of those whose fathers were not farmers (49,0%). Three decades later, not only had this balance with regards to the social origin disappeared, but also the presence of persons of non-agricultural origin became extremely dominant among non-farmers. According to the 1991 population census results, these are the percentages: 62,6% for Central Serbia and 65,5% for Vojvodina, but still high enough to illustrate the depth and lively dynamics of socio-structural changes that occurred in past thirty years.

In the current composition of labour force, the numerical domination of farmers' descendants (52,6%) over non-farmers' descendants (47,4) only occurred in Serbia Central. The large representation of farmers' descendants typical of the beginning of 60's (71,4%), was reduced by the beginning of 90's to 42,1%.

But big changes did not appear only with regard to the social origin of the active population, but also with regard to their socio-professional status in the observed time interval. Some thirty years ago there also was a small numerical difference between the share of farmers (49,2%) and non-farmers (50,8%), while in the current labour force structure the difference is very conspicuous: 29,9% of farmers as compared to 70,1% of non-farmers in Central Serbia and 19,2% as compared to 80,8% in Vojvodina.

From this review of inter-generation self-regeneration of the active agricultural population on the basis of the 1991 population census data could be obtained only in general terms, that is, the "averaged" picture about this phenomenon could be obtained in the area of Central Serbia and Vojvodina. However, when we look through the prism of some individual indicators, such as sex, age and educational qualifications, inter-generation socio-professional self-regeneration of agricultural as well as of non-agricultural active population is shown as a very selective social process. In short, in the self-regeneration of home socio-professional group the share of female persons of agricultural origin, of older age and lower educational level is clearly represented. In that connection, census data, also, unambiguously show the permanent and significant decrease level of self-regeneration of the agricultural layer, that is, increase of social mobility of active persons who by birth belong to that layer. This is directly connected to accelerated processes of the country's industrialization and urbanization.

Agricultural Transformation, Rural Development and Economic Accounts of the Agriculture: Case of Niger

Aboubacar Djimrao Abdoulaye
Direction de la Statistique et des Comptes Nationaux, DSCN - BP 862 Niamey, Niger,
Tél: + 227 72 35 60
e-mail: adjimrao@caramail.com

It is in the agricultural sector that the battle for long term economic development will be won.

1. The main burden of development and employment creation will have to be borne by the part of economy in which agriculture is the predominant activity: that is, the rural sector.

If immigration to the cities in Africa, Asia and Latin America is proceeding at historically unprecedented rates, a large part of explanation can be found in the economic stagnation of the the outlying rural areas. This is where over 2.5 billion people in the third world grind out a meager and often inadequate existence in agricultural pursuits. In Africa the ratios are much higher with almost every country having rural dwellers in excess of three-quarters of the total population. In spite of the massive migration to the cities, the absolute population increase in rural ares of most third word nations will continue to be greater than that of urban areas for at least the next decade.

Traditionally, the role of agriculture in economic development has been viewed as largely passive and supportive. Based on the historical experience of western countries, economic development was seen as requiring a rapid structural transformation of the economy from one predominantly focussed on agricultural activities to a more complex modern industrial and service society.

As a result, agriculture's primary role was to provide sufficient low-priced food and manpower to the expanding industrial economy which was thought to be the dynamic, "leading sector" in any overall strategy of economic development.

2. Subsistence agriculture and extensive cultivation in Africa

As in Asia and latin America, subsistence agriculture on small plots of land is the way of life for vast majority of African people. However, the organization and structure of the African agricultural system differs markedly from those found in contempory Asia or Latin America. Except in former large sugar, cocoa, and coffee plantations of East and West Africa, the great majority of farm families in tropical Africa still plan their output primarily for their own subsistence. Since the basic variable input in African agriculture is family and village labor, the African agricultural system is dominated by three major characteristics:

a) The importance of subsistence farming in the village community.

b) The existence of land in exess of immediate requirement which permits a general practice of shiffing cultivation. This diminishes the value of land ownership as an instrument of economic and political power, and

c) The rights of each family (both nuclear and extended) in a village to have access to land and water in the immediate territorial vicinity, excluding from such access those families that do not belong to the community even though they may be of the same tribe. To conclude, we say that thirdworld nations constitute "many parts" of the global organism.The nature and character of their future development, therefore, should be a major concern of all nations irrespective of their political, ideological, agricultural or economic orientatrions. In the latter part of the twenty-first century, there can no longer be two futures. One for the few rich and the other for the very many poor. In the words of poet, "there will be only one future, or more at all".\.

Seeding Rates and Waste of Supply on TCF Report

Joo-Hwan Kim
Department of Statistics and Information Science, Dongguk University
707 Seokjang-dong, Kyongju-shi, Kyongsangbuk-do, Rep. of Korea
e-mail : jhk@mail.dongguk.ac.kr

1. Introduction

In this paper we study and review the seeding rates and waste of supply in detail included in the electronic publication "Technical Conversion Factors (TCF)" prepared by the Statistics Division and posted on the FAO web site. It is required to study and to comment on the seeding rates and waste of supply on the publication for Asian and Middle East countries.

We first check the proposal values of seeding rates and waste of supply on TCF reported from each country in Asia and Middle East. Then, we recommend more consistent and reliable data for the seeding rates and waste of supply of agricultural crops with a clear explanation and recommended references. With recommendation data, we find some statistical characteristics of the data, and provide some guideline for the research on seeding rates and waste of supply.

2. Source Data

The data shown in the TCF publication contain the information received in response to a questionnaire sent to member nations.. Variables in TCF are seeding rates, waste of supply, extraction rates, birth rates, take-off rate, live weight, etc. All data relate to national annual averages for the five-year period, 1992-1996.

The definitions of the seeding rates and waste of supply can be found in Crop Statistics– the Concepts, Definitions and Classifications (2000), FAO Statistics Division.

To study the seeding rates and waste of supply on TCF report in detail, we first need to have accurate figures. To check the numbers, we will use three ways: FAO Statistics DB, past report on TCF, and expert's opinion. FAO has a statistics database where basic and derived agricultural statistics are collected. There is a report on TCF published by FAO; Food Balance Sheets(1977).

After data screening for seeding rate and waste of supply has been completed, we will study the variables in detail. We will try to figure out what basic statistical information is involved in the data. From the results, we then present some guidelines for the data. We will concentrate on the extreme value of the seeding rate and waste of supply. We try to explain why the value of variable appears for the various countries. Based on the above process, we will provide which information is necessary and sufficient to get correct data for seeding rate and waste of supply.

3. Screen Procedure

First, we will check the proposal value of seeding rate and waste of supply reported for each country in the TCF report using the past data for the average of 1972-1977 and the computed values from FAO DB if data is available. Secondly, we will provide some statistical information (max, min, average) of the seeding rate and waste of supply.

4. Findings and Recommendations

Proposal values for each country are not the same as the computed value from FAO DB. Computed value is calculated from the average of 1992-1996. But some countries look like the recent year's data or seem to refer to a different time period than reported. So, if computed value is available and reasonable, it is better to use it.
The following two tables represent the summary statistics of the recommendation values for seeding rate and waste of supply of cereal commodities.

Table 1 *: Seeding Rates for Cereals (Recommendation value for 1992-1996)*

	WHEAT	RICE PADDY	BARLEY	MAIZE	RYE	OAT	MILLET	SORG HUM	BUCK WHEAT	CEREAL NES
Max	287	300	232	407	270	256	106	60	450	256
Min	81	29	35	15	56	40	10	10	31	20
Average(Kg/HA)	150	104	114	63	147	135	26	27	124	90

Table 2 : *Waste of supply for cereals (Recommendation value for 1992-1996)*

	WHEAT	RICE PADDY	BARLEY	MAIZE	RYE	OAT	MILLET	SORG HUM	BUCK WHEAT	CEREAL NES
Μαξ	16,5	12,7	16,0	20,0	14,3	61,0	20,0	10,0	4,1	9,0
Μιν	0,2	0,1	0,1	0,0	1,0	2,0	2,0	0,5	2,1	2,0
Αϖεραγε(%)	4,8	3,7	4,8	4,6	5,4	10,5	5,4	4,3	3,1	4,6

5. Summary

Taking into account the findings explained in the previous sections, the following suggestions can be made. 1) There are some structural problems in the TCF report. 2) We need data consistency: some figures come from other sources. There is a need to provide the sources of the data. 3) There were many mistakes in the data from many countries. Countries should strive to collect and report their figures in conformity with the guidelines and recommendations given by FAO, as regards concepts, definitions and coverage of the data. 4) The incompleteness and inaccuracy of the basic data are the major problem encountered in developing countries.

References

Technical Conversion Factors for Agricultural Commodities(2000), FAO..
Crop Statistics – the Concepts, Definitions and Classifications(2000), FAO.

Status of Food and Agriculture Statistics 1984-1985 in Countries in Asia and the Pacific Region(1985), FAO, Regional Office for Asia and the Pacific, Bangkok, Thailand

Food Balance Sheets 1975-77 Average and Per Caput Food Supplies 1961-65 Average 1967 to 1977(1977), FAO.

Acknowledgements

The author would like to thank Statistics Division, FAO. This paper reports part of the research done as a FAO consultant. The view expressed are attributable to the author and do not necessary reflect those of FAO.

Role of Household Budget Surveys in Production of Agricultural Statistics in Africa[1]

Ben Kiregyera
Uganda Bureau of Statistics,P.O. Box, 23045, Kampala, Uganda
Tel: 256 (41) 234219, e-mail: phidam@africaonline.co.ug

Abstract*:* Household Budget Surveys are presented as alternative sources of agricultural data. Data from these surveys can also be used to validate data from agricultural censuses and surveys. For horticultural food crops, household budget surveys may well be the best source of data.

1. Introduction

The traditional sources of agricultural statistics in Africa are: reporting systems (administrative records), current agricultural surveys and agricultural censuses. Under the **Reporting System**, agricultural data are routinely collected and summarised by extension staff of the Ministry of Agriculture and then collated, summarized and dispatched to successively higher levels of administration. Usually the data are not collected for statistical purposes and they tend to be subjective and inaccurate. **Agricultural Surveys** have become the main source of accurate and current agricultural statistics. These surveys are carried out annually or more frequently, often in a number of rounds. In many countries, area and yield are measured in these surveys objectively using the FAO recommended methods. Some of the main limitations of these surveys include high costs of data collection and the inability to provide small area statistics (a serious limitation given the increased decentralized governance in Africa). The **Census of Agriculture** remains the main source of data on the structure and organization of the agricultural sector. Because of high costs and complexity of organizing and conducting the census, it is invariably carried out on a sample enumeration basis and moreover, at ten-year or longer time intervals as is usually the case.

2. Role of Household Budget Surveys

Household budget surveys (HBSs) are carried out generally to: (a) provide estimates on the level and distribution of household expenses, (b) provide the basic information needed to revise the consumption basket and weights for the consumer price index, (c) provide data to improve estimates of household final consumption expenditure component of the GDP through expenditure approach, and (d) provide as a by-product, indicative data on some of

[1] A Contributed paper presented at the Conference on Agriculture and Environmental Statistical Applications (CAESAR), Palazzo dei Congressi, Rome, 5-7 June, 2001

the socio-economic aspects of the households and activities of household members. In these surveys, data are collected on daily household consumption of food items and household characteristics of members of households. Data are mainly collected using some combination of the method of interview, diary and direct measurement of food items consumed.

With appropriate assumptions, consumption data from HBSs can be used to estimate food crop production. In Uganda, for instance, current agricultural surveys stopped in late 1970s. Until the surveys are resumed, estimates of agricultural production are partly based on projections of data collected in the 1989-90 HBS. HBSs can also be used to validate data from agricultural censuses and surveys. The following table shows that there was no significant difference in the estimates of main food crops in Tanzania made by the Ministry of Agriculture and food consumption estimates from the 1992 HBS carried out by the National Bureau of Statistics.

Comparison of Ministry of Agriculture data on main food crops and HBS data on consumption in Tanzania.

Food Crop	**HBS Consumption Estimates ('000 tonnes)**	**Ministry of Agriculture Production Estimates ('000 tonnes)**
Maize	2,409	2,226
Paddy	627	392
Wheat	42	64
Millet/sorghum	559	850
Cassava	1,621	1,778
Beans	403	312
Total	**5,661**	**5,622**

Source: *A report on examination of agricultural data and economic accounts for food and agriculture in Tanzania by B. Kiregyera and R.P. Katyal et al, June 1999*

Data on horticultural crops, viz. vegetables, fruits, etc. are not collected by the national agricultural statistical systems on a nation-wide basis mainly because of technical and operational problems of collecting data on these crops. Yet, the horticultural sub-sector of agriculture is becoming increasingly important in Africa. For instance, it was estimated from the 1992 HBS in Tanzania that consumption of horticultural crops constituted about 23% of total food consumption in the country. By not estimating horticultural production, food production is under-estimated by agricultural censuses and surveys. HBSs can help to fill the data gap.

Household Budget Surveys, however, have the problem that they are cost and time-intensive, and are carried out after long periods of time, usually after 5 years.

References

B. Kiregyera and D.K. Dutta Roy: An alternative model for the estimation of food crop production (with special reference to an African situation), Sankhya*: The Indian Journal of Statistics, Vol. 42*, Series C, Pts and 2, 1980 .

Department of Statistics (Uganda): Report on the Uganda National Household Budget Survey (1998-90)*,* Entebbe, 1991.

7 June 2001

Contributed Paper Session

SURVEY DESIGN AND ESTIMATION ISSUES

Chair: D. Berze

Systematic Circular Aerial Samples of Agricultural Importance in Synodic Time Scale

Sujata Mukherjee[1]
B14/168; Kalyani - 741 235; West Bengal; India
e-mail: sujmuk@hotmail.com

D R Reynolds
Radar Ent. Unit; NRI, Univ. of Greenwich; UK.
e-mail: d.reynolds@greenwich.ac.uk

Abstract: Aerial samples of mobile data counts examined on Synodic time scale,part I (*STAT'2000, Aug 2000,Poland,abs.*) & part II (*Stab. Prob. of Stoch. Models,Jan-Feb, 2001,Hungary,abs.*) focuses attention on lunar phases of this <u>intrinsic</u> scale (*5th BS/IMS,May,2000,Mexico*};as the key factor <u>influencing</u> the activity of noctuids, similar to ground samples, the <u>insight</u> of which was proven earlier (*4th SSC/DST,WB, India*).

Keywords: aerial samples, mobile data count, lunar phases, Synodic time scale.

1.DATA

The aim of this paper is to establish a relationship between the aerial samples and lunar phases with respect to the ground samples of noctuids. The data used in this study is observational and not from a controlled experiment exclusively done for this purpose. For the purpose of this study a systematic circular sampling design was constructed based on the sample *z at 150m height*, where the $z\,(s_1)$'s are the samples taken on day i;

$$Z = \big(z\,(s_1), z\,(s_2)\,, z\,(s_3)\,, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\, z\,(s_n)\big)' \ldots\ldots\ldots\ldots(1.1)$$

at Haringhata,$23^0$5'N & 89^0E (*Riley,1995a,Eu.J.of Ent.92,639-653* & Conference *WB-AST/DGHC,1996,Mirik,abs*)[2].The important variable for analysis is the event of sampling against time in 24 hours, occurring at the same relative time, one event during midday and 4 events between dusk to dawn through pre midnight and post midnight. The midday sampling event was taken as the center of the sequence (*Cochran, 1977*).

2.Circular samples in Synodic Scale

The mathematical series [y_{tpi}], in Synodic time scale, part I, is given as;

$$y_{(L/2-14)}\ldots\ldots y_{(L/2)}\ldots\ldots\ldots y_{(L/2+14)}, y_{2(L/2-14)}, y_{2(L/2-13)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.1)$$

where $y_{(L/2-14)}$ to $y_{(L/2+14)}$ are 29 chained variables in increasing and decreasing order of lunar phases tpi, from no moon to new moon, *L/2* standing for the full moon. To establish the influence of this time scale on all species, a data oriented approach is adopted. From the data as shown in equation (1.1) only those $z\ (s_I)$'s were chosen which were required to construct the circular samples $[Y]_{(8x5)}$ for 8 lunar phases in Nov. 1992; at day=1,dusk=2, premidnight=3, postmidnight=4 & dawn=5 with 4285 mobile data counts and $[Y]_{(6x6)}$ for 6 lunar phases in March 1994, at day=1,dusk=2, premidnight=3,midnight=4,postmidnight=5 & dawn=6 with 1229 mobile data counts of 69 species + 13 unidentified ones (*Reynolds,1999, In J Pest Managem.45(3),195-205*).

Table 1: *Nov.1992 samples showing continuous 9, 8 events at i=2, 6 near last quarter*

Lunar phases	Variables	Systematic Circular Samples @ 5 per 24 hours
20ph	$[\ y_{1(L/2+5)}\ ]_{(1x3)}$	START**day → dusk → premidnight
21ph	$[\ y_{1(L/2+6)}\ ]_{(2x3)}$	……………………… day → dusk → premidnight
22ph	$[\ y_{1(L/2+7)}\ ]_{(3x5)}$ →	24 hours continuous sampling till next →
23ph	$[\ y_{1(L/2+8)}\ ]_{(4x4)}$) →	postmidnight, ………day → dusk → premidnight
24ph	$[\ y_{1(L/2+9)}\ ]_{(5x2)}$	………………………………… dusk → premidnight
25ph	$[\ y_{1(L/2+10)}\ ]_{(6x2)}$	………………………………… dusk → premidnight
26ph	$[\ y_{1(L/2+11)}\ ]_{(7x4)}$) →	postmidnight→ dawn…….. dusk → premidnight
27ph	$[\ y_{1(L/2+12)}\ ]_{(8x2)}$ →	postmidnight→ dawn**END

Table 2: *March 1994 samples showing continuous 10 events at i=3 for the last quarter*

phases	Variables	Systematic Circular Samples @ 6 per 24 hours
26ph	$[\ y_{1(L/2+11)}\ ]_{(1x4)}$	START**day → dusk→ premidnight→ midnight
27ph	$[\ y_{1(L/2+12)}\ ]_{(2x3)}$	………………………… dusk→ premidnight→ midnight
28ph	$[\ y_{1(L/2+13)}\ ]_{(3x4)}$) →	postmidnight, ….day → dusk ……..……..→ midnight
29ph	$[\ y_{1(L/2+14)}\ ]_{(4x6)}$ →	with 24 hours continuous sampling till next →
01ph	$[\ y_{2(L/2-14)}\ ]_{(5x3)}$	……………………………day………… premidnight→ midnight
02ph	$[\ y\ 2_{(L/2-13)}\ ]_{(6x2)}$ →	postmidnight→ dawn**END

3. Results and Discussions

The last quarter systematic samples in 1992 show slight discrepancy by falling on the borderline of 3rd/4th quarter of the Synodic month. Depending on mobile data counts, aerial sampling events where skipped or continued. Also for external perturbation or some other reason the systematic sampling was disturbed in both years. For 1992, the sample size (n=8) is even, unit (k=5) is odd, $n \neq k$ and $N=25 \neq nk$; whereas for 1994 (n=6) and (k=6) both are even, $n = k$ but $N = 22 \neq nk$. Hence the disturbances are expected to be negligible because the performance of the systematic sampling for natural population is said to be precise for heterogeneous units. The purpose of this study was only to look at the time frame for proving aerial samples as efficient as ground samples (*paper forwarded, SPA'27, July, 2001,Cambridge*). The likely reason for discrepancy, if any, is the accumulated flight time, which is less than the sampling time, especially for noctuids. The perfection of the result for 1994 aerial samples,similar to ground samples (*Bowden et al 1988, Indian J. Agr.Sciences,58,2,125-130*) further indicate lunar phases, as the key factor for population growth in natural situation.This simple result lead for generalizing **HMM & ARIMA models in Synodic scale** (*papers accepted for 3rd BNP, July-August, 2001, Michigan & 23rd EMS,August,2001,Funchal*).

(1) The paper is financially supported by none till date.
(2) The present data events are the same.

Estimating the Current Mean in Successive Sampling Using a Product Estimate

Artés Rodríguez, Eva M.

García Luengo, Amelia V.
Departement of Statistics and Applied Mathematics,University of Almería, Spain
e-mail: eartes@ual.es & amgarcia@ual.es

Abstract: We study the problem of estimating a population mean, based on samples of a different size and selected over two occasions. We present a double-sampling product estimate for a matched portion of the sample if the auxiliary variable is negatively correlated with the main variable. We get the expressions for the optimum estimate as its variance and we compute the gain in efficiency of the combined estimate over the direct estimate when the information gathered in the first occasion has not been used.

Keywords: Product estimator, successive sampling, gain in precision.

1. Introduction

Successive sampling has been extensively used in applied sciences and the environment to provide more efficient estimates of population characteristics such as means. The problem of sampling on two successive occasions with a partial replacement of sampling units was first considered by Jessen (1942) in the analysis of a survey which collected farm data. In a later paper (Sen, 1971), this sampling plan was applied with success in designing a mail survey in Ontario of waterfowl hunters who hunted successively during 1967-68 and 1968-69. An estimate was developed for the current season (1968-69), based on the relationship between the value of a characteristic during the current season and its value during the previous season, which yielded more precise estimates of the kill of waterfowl than the usual estimates based on simple random sampling and using the hunter's current season's performance only. Frequently, the study of environmental issues involves negatively correlated characteristics. So, the product method of estimation is relevant to these cases. Because in many agricultural surveys computations involving product estimates become relatively complex, we propose to investigate in this paper some theory of successive sampling using a product estimate.

2. The product method of estimation

Let a simple random sample of size n' be selected on the first occasion from a universe of size N. Let a simple random sample of size m be subsampled from the n' units and let a simple random sample of size u (unmatched portion) be taken on the second occasion from the universe N-m left after omitting the m units. Also let the total number of those

sampled on the second occasion be $m + u = n$. We assume that simple random sampling is used and the finite population correction factor is ignored.
The unmatched and matched portions of the second occasion sample provide independent estimates of the population mean on the second occasion. For the matched portion an improved estimate of the population mean may be obtained using a double sampling product estimate. Hence, in table 1 we obtain minimum variance of the combined estimator. Also the optimum matching fraction, obtained minimizing in the expression of the variance with respect to u, and the gain in precision of the combined estimate over the usual estimate are given in table 1

Table 1: *Expressions for minimum variance, optimum matching fraction and gain.*

$V_{min}\left(\bar{y}_{2p}\right) = \frac{S_y^2}{n}\frac{\theta+(\theta-p)Z}{\theta+Z(1-p)(\theta-p)};\ \theta=\frac{n'}{n},\ p=\frac{m}{n}$	$\Delta = \frac{S_x}{\bar{X}} / \frac{S_y}{\bar{Y}},\ Z=\Delta(2\rho+\Delta)$
$p_{opt} = \begin{cases} \min\left\{\frac{\theta\sqrt{1+Z}}{1+\sqrt{1+Z}},1\right\} & \text{si } Z<0 \\ 0 \ \text{ó}\ \theta & \text{si } Z>0 \end{cases}$	$G = \frac{-Z\,(\theta-p)\,p}{\theta+Z\,(\theta-p)}$

Figure 1: *Gain in precision and optimum matching fraction*

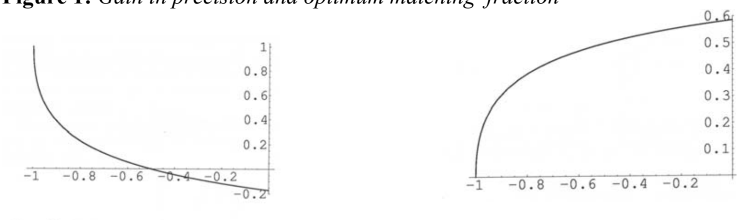


Table 2: *Gain in precision for diferrent values de ρ, θ, p (Δ=1) and ρ, Δ, p (θ =1).*

	$\theta=0.75$		$\theta=1.00$		$\theta=1.25$			$\Delta=1$		$\Delta=0.75$		$\Delta=1.25$	
$\rho\downarrow$	p_{opt}	$G_{1(max)}$	p_{opt}	$G_{1(max)}$	p_{opt}	$G_{1(max)}$	p	0.3	0.5	0.3	0.5	0.3	0.5
-0.7	32.7	9.5	43.6	12.7	54.6	15.9		11.7	12.5	15.5	16.1	4.5	5.2
-0.8	29.1	16.9	38.7	22.5	48.4	28.1		21.7	21.4	24.2	23.4	13.2	14
-0.9	23.2	28.6	30.9	38.2	38.6	47.7		38.2	33.3	36.8	32.5	27.8	26.2

References

Artés, E. y García, A. (2000) A note on successive sampling using auxiliary information, *Proceedings of the 15th International Workshop on Statistical Modelling*, 376-379.

Nonparametric Estimation of Discontinuous Spatial Functions

C.T. Jose
Central Plantation Crops Research Institute, Regional Station, Vittal 574 243, INDIA,
email:ctjos@yahoo.com

B. Ismail
Department of Statistics, Mangalore University, Mangalagangotri-574 199, INDIA,
ismailbn@yahoo.com

1. Introduction

In this paper we address commonly encountered situations in environmental/agricultural statistics or spatial/satellite images where data such as soil characters, environmental variables, incidence of pest and diseases etc. are available in a spatially divided pixels/locations. Nonparametric (bivariate smoothing) technique has been used to estimate the unknown spatial function that is assumed to generate the observations. In parametric modeling techniques, we assume some parametric form for the function and estimating the parameters based on the sample of observations. Often, there may not exist any suitable parametric form to represent the spatial function or even if it exists, we may not know the exact functional form in advance to estimate the function. In nonparametric modeling techniques, the only assumption about the form of the function is that it is smooth. In practice, we encounter the situations where there exist sudden changes in the variable and we are interested to see the location and size of sudden changes or discontinuities in the function. In the present paper, a method is proposed to estimate the jump location and jump spatial function nonparametrically. The estimated spatial function can be used to construct graphs/maps to see the changes visually. The method is illustrated through a simulation study.

2. Proposed method

The general spatial model is of the form

$$z_i=m(x_i,y_i)+\varepsilon_i \ , i=1,\ldots,n$$

where z_i is the value of the observation from the spatial location/pixel $(x_i, y_i)\in[0,1]^2$, m is the spatial function assumed to be smooth and ε_i's are i.i.d random errors with mean zero. Kernel weighted local linear regression method (Ruppert and Wand, 1994) is used to estimate the spatial function.

The estimate of m(x, y) is the solution for b_0 to the following problem:

$$\text{Minimize } \sum_{i=1}^{n}\left[z_i - b_0 - b_1(x-x_i) - b_2(y-y_i)\right]^2 K_i; \qquad K_i = K\left(\frac{x-x_i}{h_1}, \frac{y-y_i}{h_2}\right)$$

Where K is a bivariate kernel density function and h_1, h_2 are scalar bandwidths. This estimate is not suitable when there exist sudden jumps or discontinuities in the function m. The spatial regression function with a jump is defined as follows:

$$m(x,y) = g(x,y) + a(x)I_{y>c(x)}, \qquad (x,y) \in [0,1]^2$$

Where, *g(x, y)* is the continuous part, *c(x)* denotes the jump location curve, *a(x)* is the jump size function and *I* is the indicator function. The functions a, c and g are assumed to be smooth. Define the set $Q_i(x, y)$, i=1,…,4 as the set of points in the i^{th} quadrant with respect to the point (x,y). At any point (x, y), consider the following two weighted least squares (minimization) problem:

$$(1)\sum_{i=1}^{n}\{z_i - b_0 - b_1(x-x_i) - b_2(y-y_i) - a_0(x,y)I[(x_i,y_i)\in Q_1(x,y)]\}^2 I[(x_i,y_i)\in Q_1(x,y)\cup Q_3(x,y)]K_i$$

$$(2)\sum_{i=1}^{n}\{z_i - b_0 - b_1(x-x_i) - b_2(y-y_i) - a_0(x,y)I[(x_i,y_i)\in Q_2(x,y)]\}^2 I[(x_i,y_i)\in Q_2(x,y)\cup Q_4(x,y)]K_i$$

Note that when the slope of the jump location curve at any (x, y) $\in$ *c* is negative, then for a small bandwidths h_1 and h_2, the points in Q_1 (x, y) and Q_3 (x, y) will be at opposite sides of *c (.)*. Similarly, if the slope of *c* at (x, y) is positive, the points in Q_2 (x, y) and Q_4(x, y) will be at opposite sides of *c(x)*. Let the estimates of $a_0(x, y)$ obtained from the least squares problems (1) and (2) corresponding to the point (x, y) are denoted by $\hat{a}_{01}(x, y)$ and $\hat{a}_{02}(x, y)$ respectively. Between these two estimates, the estimate with maximum absolute value is denoted by $\hat{a}_0(x, y)$. The estimate of the jump location curve is given by

$$\hat{c}(x) = \underset{y\in[h_2,\, 1-h_2]}{\arg\max}\,|\hat{a}_0(x,y)|$$

Where $\hat{a}_0(x, \hat{c}(x))$ is the estimate of the jump size function *a(x)*. Jose (1999) has given the asymptotic properties of the above estimate. The spatial function *m(.)* on both sides of $\hat{c}(x)$ can be estimated separately based on the observations on either sides of $\hat{c}(x)$ by the method of kernel weighted local linear regression (Ruppert and Wand, 1994).

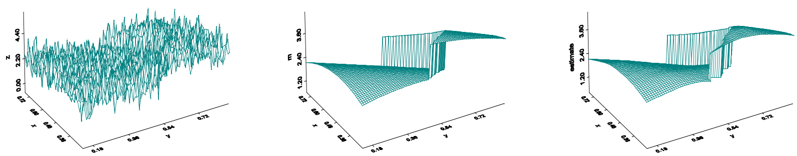
3. Simulation Study

The discontinuous spatial function considered for the simulation study is given by

$$z=1+2\sin(3x+2y)+[1+\sin(3x)]I_{y>0.7\sin(3x)}+\varepsilon$$

The error process ε is taken as $N(0, \sigma^2)$, where σ =0.80. Regularly spaced 100x100 design points (x_i, y_i), with $x_i=i/100$, $y_j=j/100$, i, j=1,…,100 have been considered for the simulation study. Based on the above, a set of data is simulated to see the performance of the proposed procedure. The kernel function considered is $K(x, y)=0.75^2(1-x^2)(1-y^2)$ and the bandwidths h_1 and h_2 are taken as 0.12. One set of simulated data, the true spatial function and the estimated spatial function based on the simulated data are given in Fig.1. It can be seen that the estimated and the true spatial functions are almost same.

Fig. 1



Simulated data True function Estimated function

References

Jose, C.T. (1999) Nonparametric inference on structural changes in regression, *Unpublished Ph.D. Thesis Submitted to Mangalore University, India*

Ruppert, D. and Wand, M.P. (1994) Multivariate locally weighted least squares regression, *Annals of Statistics*, 22, 1346-70.

Measurements of Continuous Quantities and Statistics

Reinhard Viertl
Vienna University of Technology
1040 Wien, Austria
e-mail: viertl@tuwien.ac.at

Keywords: Data Quality, Measurements, Non-precise Data, Statistics

1. Measurements

The result of a measurement of a continuous quantity is not a precise number but more or less non-precise. This kind of measurement uncertainty is different from measurement errors and stochastic uncertainty. Therefore, an adequate description of measurement results is necessary.
Examples are yields of corn from a field and amounts of poisons released to the environment.
In case of one-dimensional quantities the concept of *nonprecise numbers* is suitable. For vector valued quantities so-called *non-precise vectors* are the best up to date model to describe the measurement results.
The mathematical description of non-precise numbers and vectors will be given in the next section.

2. Non-precise Numbers and Vectors

A non-precise number x^* is defined by its so-called *characterizing function* $\xi(\cdot)$. Characterizing functions are generalizations of indicator functions of numbers or intervals. This leads to the following definition of characterizing functions.
The characterizing function $\xi(\cdot)$ of a non-precise number is a real function of a real variable obeying the following:
(1) $0 \le \xi(x) \le 1$ for all $x \in I\!R$
(2) There exists at least one $x \in I\!R$ with $\xi(x) = 1$
(3) For all $\delta \in (0,1]$ the so-called δ - cut $C_\delta[\xi(\cdot)]$, defined by

$$C_\delta[\xi(\cdot)] := \{x \in I\!R : \xi(x) \ge \delta\}, \text{ is a closed finite interval } [a_\delta, b_\delta]$$

It turns out that non-precise numbers are a suitable model for measurement results.
For k-dimensional continuous quantities non-precise observations are described by *non-precise vectors* $\underline{x}^*$ which are defined by so-called *vector-characterizing functions* $\zeta(\cdot,\cdots,\cdot)$ which are real valued functions of k real variables $x_1,\cdots,x_k$. For details see [2].

3. Non-Precise Samples and Statistics

The quantitative description of non-precise data makes it necessary to adapt statistical methods to this kind of data. This is possible using the concept of propagation of imprecision. The first step for this is the combination of non-precise observations to obtain the so-called *combined non-precise sample*.

In the simplest situation of a one-dimensional stochastic quantity X with observation space $M \subseteq I\!R$ a sample is a finite sequence $x_1^*, \cdots, x_n^*$ of non-precise numbers. These non-precise numbers have to be combined in order to obtain a non-precise vector $\underline{x}^*$ in the *sample space* M^n.

The vector-characterizing function of the combined non-precise sample is obtained from the characterizing functions of the non-precise observations in the following way: Let $\xi_i(\cdot)$ be the characterizing function of x_i^* for $i = 1(1)n$. Then the values of the vector-characterizing function $\zeta(\cdot, \cdots, \cdot)$ of the combined non-precise sample $\underline{x}^*$ are defined by

$$\zeta(x_1, \cdots, x_n) = \min\left[\xi_1(x_1), \cdots, \xi_n(x_n)\right] = \min_{i=1(1)n} \xi_i(x_i) \quad \text{for all } (x_1, \cdots, x_n) \in I\!R^n. \quad (1)$$

Based on the combined non-precise sample $\underline{x}^*$ the values of statistics $\mathcal{S}(X_1, \cdots, X_n)$ can be calculated. Let $S = \mathcal{S}(X_1, \cdots, X_n)$ be a real valued continuous function of a sample $X_1, \cdots, X_n$.

For observed non-precise concrete sample $x_1^*, \cdots, x_n^*$ the value $\mathcal{S}(x_1^*, \cdots, x_n^*)$ of the statistic S becomes non-precise, i.e. $s^* = \mathcal{S}(x_1^*, \cdots, x_1^*)$, where the characterizing function $\psi(\cdot)$ of s^* is given in the following way: Using the notation $(x_1, \cdots, x_n) = \underline{x} \in I\!R^n$ the values $\psi(s)$ of $\psi(\cdot)$ are obtained by

$$\psi(s) = \begin{cases} \sup\{\zeta(\underline{x}) : \mathcal{S}(\underline{x}) = s\} & \text{if } \mathcal{S}^{-1}(\{s\}) \neq \varnothing \\ 0 & \text{if } \mathcal{S}^{-1}(\{s\}) = \varnothing \end{cases} . \quad (2)$$

Remark: The δ-cuts of the characterizing function $\psi(\cdot)$ of the non-precise value s^* are given in the following way:

$$C_\delta[\psi(\cdot)] = \left[\min_{\underline{x} \in C_\delta[\zeta(\cdot)]} \mathcal{S}(\underline{x}), \max_{\underline{x} \in C_\delta[\zeta(\cdot)]} \mathcal{S}(\underline{x})\right] \quad \text{for all } \delta \in (0,1] \quad (3)$$

The resulting value of a statistics is not a precise number but a non-precise one. This imprecision is quantified by a characterizing function. Examples are non-precise estimations for parameters.

4. Further Statistical Procedures

All standard statistical procedures can be generalized to the situation of non-precise data. This includes classical as well as Bayesian statistical methods.
An interesting topic are non-precise a-priori densities in Bayesian inference. These make it possible to use also fuzzy a-priori knowledge in a quantitative way.
For more details see the introductory book [2] and the encyclopedia article [3] which is in the process of publication.

References

Menger K. (1951) Ensembles flous et fonctions aléatoires, *Comptes Rendus Acad. Sci.*, Paris

[1] Viertl R. (1996) *Statistical Methods for Non-Precise Data*, CRC Press, Boca Raton

[2] Viertl R. (to appear) Statistical Inference with Non-Precise Data, in: *Encyclopedia of Life Support Systems*, UNESCO

Sampling Designs in Multidimensional Populations

Nicolas Farmakis
Aristotle University of Thessaloniki, Dept. of Mathematics, GR-54006 Thessaloniki, GREECE, e-mail: farmakis@auth.gr

Abstract: Low budget designs for sampling one-dimensional subsets of multi-dimensional populations are proposed, where the elements from the selected subset of the population are drawn by one of the classical sampling techniques like the Simple Random Sampling.

Keywords: Sampling, Designs, Population, Multi-dimensional.

1. Introduction

Consider a Population **P** with elements (units) taken as points of a k-dimensional Domain D, i.e. $D \subseteq R^k$ so that every unit $\mathbf{u} \in D$ figures as the $\mathbf{u}(x_1, x_2, \ldots, x_k) \in D \subseteq R^k$, where $x_j \in R$, the j-th coordinate of **u,** j=1,2,…, k. We study the random variable H, defined in D, such that the suitable value $H(\mathbf{u}) = \mathbf{f}(x_1, x_2, \ldots, x_k) \in R$ of H corresponds to $\mathbf{u}(x_1, x_2, \ldots, x_k) \in D$. It is obvious that we work on the (k+1)-dimensional Euclidean Space, the R^{k+1}, and deal with a graph (H) related to the function **f** of k variables**.** We assume that we have some knowledge about the form of $\mathbf{f}(x_1, x_2, \ldots, x_k)$ subject -probably- to some parameters. It is well known that the mean value of H(**u**) is given by:

$$\overline{H} = \frac{1}{\|D\|} \iint\limits_{D} \ldots \int f(x_1, x_2, x_3, \ldots, x_k)\, dx_1 dx_2 dx_3 \ldots dx_k\,, \tag{1.1}$$

where $\|D\|$ is the value of the Euclidean measure (hypervolume) of D. The main goal of this paper is **the estimation of** $\overline{H}$ **,** based only on a sample of points (units of **P**) taken from an arc $T \subseteq D$ with dimension m, $1 \le m < k$. Note that in most cases m is equal to 1.

2. The Sampling Design as an algorithm

The main idea in this paper is to use analytical methods for the identification of a suitable subpopulation of dimension m<k, which will serve as the population for sampling. This subpopulation is a subset of a subspace (T) of R^k. We are going to draw elements only from the arc $T=(T) \cup D$. Note that if the form of **f** involves unknown parameters then a presampling procedure may be used to provide point or interval estimators for the unknown parameters. Naturally, the best scenario for us is when the exact value of the integral in

(1.1) is known. In any case, if the integral does not exist or it cannot be calculated by a finite number of analytical operations, then we need some help of numerical methods. The basic steps of the method for the estimation of $\overline{H}$ are:

STEP 1: Collect as much information as possible about function H(**u**)= **f**($x_1,x_2, \ldots,x_k$)/**u**∈ D.

STEP 2: Evaluate $\overline{H}$ given in (1.1) by integration, based on information of STEP 1.

STEP 3: Look at the arc T=(T)∪D and denote the measure (norm) of T, as $\|T\| \in$ R. See that arc T⊆R^m ⊆R^k and the curve (T) is a m-dimensional subspace of R^k, m<k, i.e. T is the set of points (elements):

$$T=\{(x_1,x_2,\ldots,x_k),x_{r_{m+t}}=\tau_{r_{m+t}}(x_{r_1},x_{r_2},\ldots,x_{r_m}),x_{r_i}\in I_{r_i}\subseteq R,\ i=1,\ldots,m,t=1,\ldots,k-m\} \qquad (2.1)$$

STEP 4: Evaluate the mean value of H over T, as

$$\overline{H}'=\frac{1}{\|T\|}\cdot\iint_T\ldots\int f\left(x_{r_1},x_{r_2},\ldots,x_{r_m},\tau_{r_{m+1}}(x_{r_1},\ldots,x_{r_m}),\ldots,\tau_{r_k}(x_{r_1},\ldots x_{r_m})\right)dx_{r_i}dx_{r_2}\ldots dx_{r_m}\,. \qquad (2.2)$$

STEP 5: Look at the difference $\Delta\overline{H}=\left|\overline{H}'-\overline{H}\right|$ and solve one of the next equations, in order to obtain the parameters of the curve (T):

$$\Delta\overline{H}=0,\ \ \Delta\overline{H}=\min_T\Delta\overline{H} \qquad (2.3)$$

STEP 6: Get a sample (e.g. random) drawing points from the arc T and calculate the estimator $\hat{\overline{H}}$ of $\overline{H}$ (Cochran, (1977), Ch. 2; Farmakis, (1999); Farmakis, (2000), Ch. 2).

3. An application

In general, the traditional sampling techniques are both time and money consuming, especially when we are dealing with multidimensional populations. The technique presented here, which does not suffer from the above problems, can be used, among other settings, in agriculture where the sampling from an entire field can be replaced by the sampling on a single straight path. These comments as well as the next example show that the contribution of the present work lies on the fact that the proposed technique is extremely useful due to both its simplicity and its low budget.

Example: Take k=2, m=1, H=f(x,y)=axy, and D={(x,y), $0\le x,y\le 1$}. The proposed method gives as a solution the line y=0.75x and $\Delta\overline{H}=0$, in T={(x, 0.75x), $0\le x\le 1$}, with $\overline{H}=0.25a$. A random sample of 16 values of x, obtained by the CASIO fx-85VH pocket calculator, is: {.2, .968, .531, .276, .256, .641, .639, .464, .586, .137, .514, .755, .868, .066, .413, .973}. Simple calculations show that $\hat{\overline{H}}$ =0.2581246a$\cong 0.25\cdot a$.

References

Cochran, W.G. (1977) *Sampling Techniques,* Wiley, New York.

Farmakis, N. (1999) Systematic Sampling in Two-dimensional Populations, Proc. of the 52nd International Congress of ISI, Contr. Papers B.-1, pp 319-320, Helsinki, Finland.

Farmakis, N. (2000) *Introduction to Sampling.* Christodoulidis Publ/ons, Thessaloniki (in Greek).

Area Estimators Variance from Systematic Samples

Javier Gallego
JRC, I-21020 Ispra. Italy,e-mail: javier.gallego@jrc.it

Abstract. We consider two-dimensional systematic samples on a square grid for an example using CORINE Land Cover as pseudo-ground-truth. A variance estimator based on the comparison of each observation with the neighbouring values is compared with the real variance and with the result of applying the usual variance estimator for random sampling.

Keywords: land cover area estimation, systematic sampling. area frame sampling.

1. Two-dimensional systematic sampling.

Systematic sampling is generally more efficient than simple random sampling (SRS). One-dimensional systematic sampling is optimal if the autocorrelation is positive, decreasing and convex (Bellhouse, 1988). For point sampling in the plane, triangular grids perform slightly better than square grids (Olea, 1984), but square area units are more practical for field survey.
The main drawback of systematic sampling is the absence of an unbiased estimator for the variance. The classical variance estimation formulae for random sampling are sometimes used, but this generally overestimates the variance. Other options are splitting the sample or combining several replicates (Koop, 1971), but estimates of the variance are often unstable. Drawing several replicates reduces the efficiency (Gautschi, 1957). Other variance estimators compare each sample element with neighbours. Wolter (1984) compares several estimators of this type for the one-dimensional case. Matern (1986) studies similar estimators for the plane with an assumption of stationarity, that may be debatable. We test here neighbourhood variance estimators with semi-real data.

2. Test data set and results

CORINE Land Cover (CLC) is a land cover map made by photo-interpretation of satellite images and additional information (CEC, 1993). We use it as pseudo-truth, so that we can assess different estimators knowing the whole population. We assess estimates for the area of 7 major land cover types in Andalucía (Spain).
A 1 km grid has been overlaid keeping only the cells fully inside the region to simplify boundary effects (N=86715). Each cell (i, j) is identified by a row i and a column j. All possible 100 systematic samples are drawn with a 10 km step (1% sampling rate).
The target variable is Y_c: % land cover class c. The next variances of $\bar{y}_c$ are compared:

- The expected variance with simple random sampling (SRS).
- The expected variance with systematic sampling (this can be computed because the complete population is known).
- The average estimated variance using SRS formulae on systematic samples .

- The mean estimated variance from differences between neighbouring observations:

$$\hat{V}(\bar{y}_c) = \frac{1}{2m\,n}\sum_{ij} d_{ijN}^2 + d_{ijE}^2 \qquad \text{where} \quad d_{ijN} = y_c(i,j) - y_c(i+10,j) \qquad (1)$$

d_{ijE} is the difference with the neighbour to the east, and *m* is the total number of pairs.

Table 1: *Variance estimators of land cover from CLC in Andalucía*

	area (km^2)	SRS Std error	Systematic Std. Error	Systematic Relat. Eff.	Estimated std. error as SRS	% bias variance	from neighbours	% bias variance
Artificial	1165	235	204	1.32	236	34	220	16
Arable	24626	1125	750	2.25	1131	128	826	21
Perm. Crops	12242	844	646	1.71	848	73	654	3
Heterog. Agric.	12859	867	660	1.72	872	74	736	24
Forest	11555	811	506	2.57	816	160	691	87
Nat. veg	22933	1053	943	1.25	1058	26	876	-14
Water	1336	300	161	3.47	301	250	231	106

Table 1 gives the values of standard errors of the area estimates (in km^2), which is easier to interpret than the variances of $\bar{y}_c$. Relative efficiency and bias are computed from the variances. Several comments can be made:

- Systematic sampling is more efficient than SRS, confirming results in the literature.
- The usual SRS variance estimator strongly overestimates the variance of systematic sampling. The estimated variance appears to be slightly worse than with SRS.
- Estimating the variance by comparing neighbouring values generally gives a moderate overestimation. The correlograms for certain land cover types still need to be analysed to understand why a strong overestimation or underestimation occur in these cases.

References

Bellhouse D.R. (1988), Systematic sampling, in *Handbook of Statistics*, ed. P.R. Krisnaiah, C.R. Rao, pp. 125-146, North Holland-Elsevier, Amsterdam

CEC (1993), *CORINE Land Cover; guide technique*, Report EUR 12585EN. Office for Publications of the European Communities. Luxembourg,. 144 pp

Gautschi W. (1957), Some remarks on systematic sampling, *Annals of Mathematical Statistics*, Vol. 28, pp. 385-394

Koop J.C. (1971), On splitting a systematic sample for variance estimation. *Annals of Mathematical Statistics*. Vol 42, n° 3, 1084-1087

Matern B. (1986), *Spatial variation*. Springer Verlag lecture notes in statistics, n. 36.

Olea R.A. (1984), Sampling design optimization for spatial functions, Mathematical Geology, 16, 4, 369-392.

Wolter K.M. (1984), An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association*, Vol. 79 No 388, pp. 781-790.

Satellite Meeting

4 June 2001

SATELLITE MEETING

Organised by IWG.AGRI and ISTAT

Agricultural Statistics in the CIS Countries (General Overview)[1]

Youri Ivanov
Interstate Statistical Committee of the CIS, Moscow, 39, Myasnitskaya Str., Russian Federation
e-mail: youri.ivanov@cisstat.com

Keywords: agricultural statistics, output, intermediate input, survey, census.

1. Introduction

The purpose of this paper is to review briefly the current situation in agricultural statistics of the CIS countries and to identify the problems which require attention and need to be dealt with in the foreseeable future in order to improve the coverage, quality and international comparability of this statistics.

As will be shown in detail below, statistical offices of the CIS countries produce and publish a variety of agricultural statistics. Unfortunately this statistics is quantity-oriented and there are serious gaps in this information which inhibit serious analysis of many qualitative aspects of agricultural economy essential for making decisions on formulation of agricultural policy. A bulk of the indicators computed by the statistical offices have been inherited from the USSR times and a few new indicators such as value added on the basis of the SNA93 are not sufficiently harmonized with the old indicators. Although the published figures on production of major crop and animal products are, by and large, in line with the FAO definitions, in some cases deviations from the latter are significant. Though sample surveys are employed on a larger scale as compared with the practice which existed in the USSR, there are still visible shortcomings in the process of collection of primary data and many FAO recommendations in this area are not fully implemented; thus, agricultural censuses and censuses of cattle, in particular, are not carried out as frequently as the FAO suggests.

The review is intended to focus on such topics of the agricultural statistics as economic accounts; crop and animal commodity production statistics; trade and prices statistics; labor and capital input statistics; income statistics; farm registers; agricultural surveys and censuses.

The review deals with the availability of the data pertaining to the above topics, underlying definitions and classifications, major sources of data, consistency with the international standards, degree of accuracy of data in a broad sense, encountered problems.

[1] The present paper is financially supported by the Statistical Office of the ECE.

2. Economic accounts

It is worth noting that in the not so distant past economic accounts for agriculture were compiled on a large scale both by the statistical office of the USSR and the statistical offices of the individual republics. These accounts were compiled for an impressive number of agricultural commodities (or groupings of commodities) both in physical units and in monetary form; they were compiled for major types of agricultural enterprises and holdings: state enterprises, collective farms, personal plots of employees, personal plots of collective farmers. These accounts yielded a number of data essential both for analysis of the agriculture and for compilation of national accounts (so called the Material product system, also known as the Balance of National Economy). These data referred to output, sales, intermediate and final consumption, value added, fixed capital formation (increase in stocks of cattle), changes in stocks of goods, etc. Unfortunately this system of accounts (which was also used in many other centrally planned economies and was described in detail in some documents of the European Conference of Statisticians) was practically dismantled during the early years of transition to market economy in practically all CIS countries (largely due to the lack of resources) and only few fragments of this system were retained.
Some of these fragments were transformed in order to meet the requirements of the SNA 93. At the present time all the CIS countries compile production and generation of income accounts for industries (and for agriculture, in particular) and many CIS countries have started compiling national accounts for the institutional sectors where agricultural enterprises (both corporations and unincorporated entities) are represented. In the context of this work the countries compile the following items of information on agricultural economy:

- output in current prices (basic values);
- intermediate consumption which covers input of agricultural and industrial goods and payments for services;
- gross value added in basic prices;
- compensation of employees;
- other taxes on production, net of subsidies;
- gross operating surplus / gross mixed income.

In addition some pieces of data on disposition of agricultural output are computed in order to provide the items of information needed for compilation of other accounts of the System; thus, the data on final consumption of agricultural goods, including data on own final consumption of agricultural goods, gross fixed capital formation (change in stocks of cattle), change in inventories and external trade flows are estimated to provide information needed for compilation of “goods and services account“ of the SNA. In many cases these figures are obtained relatively independently from the estimates of output and this may lead to inconsistencies and discrepancies between the figures on sources and disposition of agricultural goods. A reconciliation of the figures on resources and disposition of agricultural goods is achieved in the aggregated input-output tables which are periodically compiled in the CIS countries and in which agriculture is normally represented as one branch or two sub-branches: crop products and animal products. However, since this reconciliation is carried out on a very aggregated level, it is difficult to eliminate the discrepancies at the lower level of aggregation (because they are likely to cancel each other out in the process of reconciliation at the highest level of aggregation). The results of this reconciliation can be improved if the supply and use tables (or input – output tables) are compiled with a sufficient degree of disaggregation which would enable us to distinguish between major agricultural commodities (or groupings of commodities). It appears, however, that most of the CIS countries will not be able to carry out such detailed compilations in the foreseeable future due to lack of resources.

Some CIS countries have undertaken special efforts to enhance consistency between the various types of agricultural statistics. Thus, the Russian Goskomstat has introduced recently, for agricultural enterprises, a monthly integrated reporting form (П – 1(сх), entitled «The data on production and delivery of agricultural products» which contains integrated information on output and sale of major crop and animal products, stocks of cattle and agricultural goods, stocks of fodder, and on the stage of sowing or harvesting work. This form (though it is compiled only in physical units) seems to be a step in the right direction. However it can not resolve all the problems of harmonisation of agricultural statistics.

It is worth noting that there are two methods of computation of agricultural output which have been recommended by the CIS statistical Committee. In accordance with the first method the output of agriculture is taken to be equal to the sum of outputs of agricultural establishments (the latter include not only establishments of agricultural enterprises but also the agricultural establishments of non-agricultural enterprises).

However, in practice the CIS countries use the second method which is also recommended by the CIS Statistical Committee. This method is normally referred to as gross turnover method. In accordance with this method the output of agriculture is defined to include the total value of all agricultural products irrespective of where they have been produced: in agricultural enterprises, personal plots of households, in agricultural establishments of non-agricultural enterprises or produced as a result of the secondary activities of non-agricultural enterprises. On the other hand, non-agricultural goods produced i) as a secondary output of agricultural establishments and ii) by non-agricultural establishments of agricultural enterprises are excluded from the output of agriculture. The important implication of this method is that the output includes all types of own intermediate consumption (all seeds and fodder produced and consumed in the given year are included irrespective of where they were used for production purposes).

More specifically, the output of agriculture as measured in accordance with the gross turnover method includes the following items:

- sales and barter;
- goods supplied to non-agricultural establishments of the same agricultural enterprises (e.g., milk supplied to the children institution owned by the enterprise, or goods delivered to other establishments of the same enterprise for processing);
- wages in kind;
- own final consumption in holdings of households;
- own intermediate consumption;
- change in stocks of agricultural goods (produced but not sold);
- change in stocks of cattle;
- change in stocks of work in progress;

- change in stocks of plantations.

It is worth noting that the total output in current prices is computed by multiplying the quantities of the individual products by the average current market prices rather than by summing up the values of the components of output valued at corresponding market prices .Therefore this procedure may result in some discrepancies between the measure of output and its disposition. In Russia commencing in the year 2000, agricultural enterprises (other than unincorporated enterprises owned by households) are supposed to submit to the statistical office on a quarterly basis two special reporting forms which contain data on value of output and sales in current prices as well as on the cost of production (forms П-1 and 5-з respectively), however, these data require some adaptation to be included in the national accounts; for the time being they are used largely for checking purposes because they do not cover either all output/costs or all producers. Nevertheless introduction of the above forms is a step in a right direction.
The figures on output of agriculture, computed by the CIS countries also include the value of selected agricultural services provided by specialized organization (e.g., veterinary services, irrigation services, agro-chemical services and so forth). These services can be market and non-market; the former are allocated to intermediate consumption whereas the output of the latter is treated as final consumption expenditure of government. The differences in organization and financing these activities in different countries may therefore affect the international comparability of value added. Changes in the organization of these activities in one country over time may affect the comparability of value added estimates over time.
It is important to note that output of agriculture includes in principle not only output of the informal sector (unincorporated enterprises which sell most part of their output as well as output unincorporated enterprises owned by households which consume most of their produce) but also some estimates of the underground economy or underreported output by the respondents. The adjustments which are introduced to cover this hidden output often rely on a rather crude methods and cannot normally claim a high degree of accuracy. That is why reconciliation of resources and uses of agricultural products should be made at the lowest possible level of aggregation, which is not the case in practice, as was mentioned above.
The output is valued at basic prices as they are defined in the SNA 93. This implies, among other things, that output is valued at prices which exist at the moment of production and not at the moment of sale of goods; in other words, efforts are normally undertaken to remove the estimate of holding gain from the output. The same relates to valuation of intermediate consumption which has to be valued at prices which exist when goods enter the production and not at the prices which were actually paid It should be recognized, however, that estimates of holding gain in both cases (output and intermediate consumption) are often produced with the help of rather crude methods and this may affect the accuracy of the value added.
Another fragment of the former system of accounts in agriculture which is maintained in most CIS countries refers to balances of resources and disposition which are compiled in physical units for a limited number of agricultural products: grain, products of primary processing of grain, potato, vegetables, melons, watermelons and similar products, fruits, meats, milk products, eggs. The balances show opening and closing stocks of goods and the flows which explain the changes in stocks such as output, imports, intermediate and final consumption, processing and exports.

3. Production of agricultural goods in physical units, trade and prices

All the CIS countries compute and publish figures on production of major crop and animal products in physical units. The bulk of these data are collected and disseminated by the CIS statistical Committee. These data rely on similar methods (on the whole consistent with the international recommendations) and sources of primary data and therefore can claim a degree of international comparability .The list of crop and animal products the figures on output of which are published includes:

Grain	Flax	Citrus fruit
wheat	Raw cotton	Grape
rye	Sugar beets	Milk
oats	Potato	Eggs
barley	Vegetables	Wool
corn	Melons	Meat (by type)
leguminous crop	Fruits and berries	Output of growing cattle

The statistical offices of the CIS countries compile figures on output of some other crop and animal products such as rice, sunflower seeds, mustard, soy, hay, millet, some material for plantations, honey, skins and some other products but the data on production of these commodities are not collected by the CIS Statistical Committee.
The above data on production of individual crop and animal products are used to derive output of agriculture in constant prices and compute volume indices.
As was mentioned above, the underlying definitions and methods of measurement are, on the whole, consistent with the international recommendations and yet there are some deviations which are worth mentioning. Thus, FAO classification of grain is much more detailed than used by the CIS countries (up to 13 brands); furthermore, FAO recommends to register separately different brands of wheat and corn (so called hard brands) but this is not systematically done in the statistics of the CIS countries either. Output of early harvest of potato is not shown separately as the FAO suggests. The recommendation of the FAO to register output of potato after removing pieces of soil and dirt is not implemented either. Prior to 1998 some deviations from the FAO recommendations existed with regard to the methods of calculation of the yield per unit of land. Thus, the CIS countries relied in the context of this analysis on the land which included the land where crop perished or where the crop was not harvested for one reason or another. Commencing in 1998, the countries use the land actually harvested for calculation of the yield per unit of land. This was a result of implementation of the material “Recommendations on computation of the yield per unit in accordance with the FAO methodology” prepared by the CIS Statistical Committee for the CIS countries. It should be noted in this context that CIS statistical committee has prepared a number of other documents intended to facilitate implementation of the FAO recommendations in the CIS countries. Nevertheless some deviations continue to exist in addition to those mentioned above. For example, output of eggs includes some losses which are to be excluded in accordance with the FAO methodology. There are some marginal differences in the scope of output of meat and wool between the CIS and FAO methodologies; for example, contrary to the FAO recommendations the output of meat published by the CIS countries include some fat and so called sub products; the latter may account for up to 3 per cent of the total output.

Statistical data on cattle is not detailed enough to meet the FAO recommendations; thus, the latter suggest to work out annual data for each brand of cattle on opening and closing stocks as well as on all types of flows which link the opening and closing stocks such as births of animals, natural losses, exports and imports and slaughtering. These types of analysis are not carried out in any CIS country even at the most aggregated level, that is for cattle taken as a whole. Measuring of output of growing cattle in monetary form requires data on slaughtering and preferably separate data on change in stocks of cattle allocated to inventories and fixed assets (the latter is especially important for correct valuation of stocks); it also requires special valuation of these elements. In practice, however, they are not always available or they are not complete or they do not meet all the requirements of standard methodology and principles of the SNA.

In addition to the figures on output of agricultural products the statistical offices of the CIS countries publish data on average yield per unit of land and productivity for a rather detailed list of crop and animal products; some countries such as Russia, for example, publish data on the average yield of grain per unit of land broken down by major types of grain (winter and spring grain, winter and spring barley, rye, oats, corn); the bulk of this information is collected and disseminated by the CIS statistical committee.

Compilation of data on production of crop and animal products relies on a variety of sources. The specialized agricultural enterprises (state, cooperative, private) submit records to the statistical authorities. For example, in Russia private farms submit to the statistical office two forms: “F.2 Data on output of crop products” and “F.3 Data on output of animal products and number of cattle”. The figures on output produced in personal plots of households, including small farms are computed by using indirect methods which rely on the established size of agricultural land (stocks of cattle) and the estimates of the average yield per unit of land (average productivity); the latter are obtained with the help of various sample surveys. The data on agricultural land are obtained from the records of agricultural enterprises and special periodic surveys and censuses; the same relates to sources of data on stocks of cattle. Complete censuses of cattle are carried out once every ten years and this is still another deviation from the FAO recommendations on this matter. The data on stocks of cattle held by small farms, personal plots of households and other unincorporated enterprises of households are established annually from the special register books compiled by local authorities; these data are revised as a result of complete censuses of cattle. However, the latter are carried out rarely as was mentioned above. The data on the yield per unit of land and productivity in personal plots of households are obtained with the help of sample surveys; the list of payers of tax for land is normally used to identify the holdings to be surveys on a sample basis The list of deviations from FAO recommendations on statistics of production is not exhausted and can be extended.

Practically all the CIS countries compile data on sales of major agricultural goods (crop and animal products) both in physical units and monetary form. The data on sales include goods supplied on barter; the data on sales are normally subdivided into two categories: i) sales to the specialized procurement organizations which purchase goods for needs of the state and ii) sales through other channels (sales at the market, sales to trade organizations and enterprises of catering and so forth). The bulk of this information is collected and disseminated by the CIS Statistical Committee. The data on the average prices of sales are also computed and published by statistical offices of the CIS countries; the prices are computed excluding value added tax, but including some subsidies and therefore they approximate basic prices as defined in the SNA; the list of agricultural goods for which average prices are compiled and published includes approximately 30-35 items and they cover up to 95 percent of total output; these data provide the basis for computation of price indices; the latter are computed separately for crop products and animal products. The Laspeyres formula is used to compute this indices; the data on sales are used as weights.

4. Labor and capital input

All the CIS countries compile and publish figures on employment in agriculture in accordance with CBNE or ISIC (rev. 3). The data include persons employed both in agricultural enterprises (state, private and collective ones), in unincorporated private farms and in personal plots of households; however, employment in personal plots of persons mostly engaged in non-agricultural activities is not included even though the output of these holdings is counted as a part of agricultural output; for example, persons working at industrial enterprises and spending some time at their dachas to grow flowers, vegetables and potato are not included in agricultural employment but their output is counted as a part of agricultural production. The data on employment in agricultural enterprises (both state and private) are obtained with the help of special reporting system (record on labor); for example, Russian Goskomstat employs for this purpose the form П-4 collected on a monthly basis and the form 1-T collected annually. The data on employment in farms (unincorporated entities) are estimated on the basis of information on a number of such farms and the data of periodic surveys which make it possible to establish the average number of persons employed at the farms. The figures on employment in personal plots of households are estimated with the help of data on size of land, stocks of cattle and labor input per unit of land (per head of cattle); the latter data are established from special surveys periodically carried out by statistical offices. It should be noted that some CIS countries introduced into regular statistical practices surveys of labor force; in the context of these surveys the persons are asked whether they were engaged in agricultural activities of personal plots of households and if yes whether this employment was secondary or primary; how many hours were spent in personal plots. This information was then used to estimate the employment in personal plots of households. It is clear from the above that statistics of employment in agriculture should be supplemented by the data on man-hours spent in agriculture in order to provide a better basis for analysis of productivity, to coordinate data on output and labor input in personal plots of households and in agriculture as a whole.
The information on capital input compiled and published by the CIS statistical offices includes:

- capital investments in agriculture both in current and constant prices as well as volume indices;
- fixed assets put into operation in the given year;
- stocks of fixed assets classified by major types of assets (buildings and structures, machinery and equipment, transport means, cattle and so forth).

It is worth noting that there is a general understanding that one of the serious problems of agricultural economy in practically all the CIS countries refers to very high degree of wear and tear of fixed assets. However, no systematic and sufficiently detailed statistical data (for example, by types of assets) are available on this topic.

Some countries produce more detailed statistics on capital input. Thus, Russian Goskomstat publishes separate figures on acquisition and disposal of fixed assets, on stocks of agricultural machinery (by major groupings), on capital investments in irrigation and so forth.

It should be noted that the above indicators of various flows and stocks in agriculture are compiled outside of the framework of national accounts and therefore their underlying definitions somewhat deviate from those recommended in the SNA 1993. This refers both to scope of flows and stocks and to the methods of their valuation. For example, the stocks of cattle are often valued at cost (rather than at market prices as suggested in the SNA), because this mode of valuation is used in business accounts of agricultural enterprises which submit their data to statistical offices. Valuation of stocks of cattle at cost creates problem with the measuring the output of growing cattle in accordance with the principles of national accounting. Contrary to the SNA 93, capital investments figures compiled in the framework of conventional agricultural statistics do not include capital repair, acquisition of cattle (classed as fixed assets), purchases of some intangible assets (e.g., software), transfer costs on purchases of non-produced assets such as land.

It should be noted in this context that the SNA 93 recommends to compile only two accounts for industries (for agriculture, in particular): production and generation of income account and therefore accumulation accounts are not compiled for industries. That is why the reconciliation of the above flows on capital input with the SNA definitions and classifications has not been yet undertaken on a systematic basis in any CIS countries.

5. Income

Data on income of agricultural enterprises and population engaged in agriculture are very limited (in the case of enterprises) or practically not available (in the cases of agricultural population) in any of the CIS countries. As a rule, data on gross income and profits of medium and large enterprises are compiled and published. However, the underlying definitions of these indicators differ from those recommended in the SNA 93 for similar flows. Data on income of other types of agricultural producers (unincorporated farms, personal plots of households are not available). As was mentioned above, the only comprehensive information on this topic which is available now in all the CIS countries refers to generation of income account compiled for the agriculture where value added created is decomposed into its major components, including profits and mixed income. This information could be used as a starting point for a more detailed income analysis in agriculture. It should be reminded again that the other accounts dealing with the income analysis are not compiled (according to the SNA 93) for industries, but only for the institutional sectors. It means that in order to compile figures on disposable income and saving in agriculture, the flows on primary income as payable and receivable by agriculture (as well as on the current transfers) should be extracted from the relevant income distribution and redistribution accounts compiled for the non-financial enterprises where these flows are registered together with similar flows payable to and receivable from other institutional units. This is a time consuming and tedious work, but some data are available for this exercise, e.g. on income from property payable and receivable by agricultural enterprises, on taxes payable by agricultural enterprises both on income and land, on insurance premiums and claims payable and receivable by agricultural enterprises and so forth.

The data on income of population engaged in agriculture are not compiled by the CIS countries and there is currently no methodology available which can be immediately used for this purpose. Some fragments of information on income of agricultural population are contained either in so called "Balance of money income and expenditure of population" (which in the past was an important element of the MPS and which is still continued to be compiled by all the CIS statistical offices) or in income accounts of household sector compiled in many CIS countries which started implementing the SNA 93. But again, a great deal of work is required to extract the data needed for analysis of income of agricultural population from the above sources.

6. Registers, surveys and censuses

In all the CIS countries the registers of enterprises and organizations belonging both to private and public sectors have been set up. In addition, statistical offices set up agricultural sub-registers which contain detail data on agricultural enterprises, unincorporated farms, personal plots of households, including personal plots of non-agricultural households, where vegetables, potato, fruits and flowers are grown. In some countries such registers contain a significant number of indicators and represent a computerized database. For example, in Russia sub-registers of agricultural enterprises contain 250 indicators, while the sub-register of unincorporated farms includes 70 indicators describing activities of this holding. A considerable number of indicators is contained in sub-registers of Kazakhstan. Completeness and reliability of such registers is achieved by statistical offices through close cooperation with the administrative bodies (which have their own registers) and with tax inspection agencies. Agricultural sub-registers are used as a basis for organization of various statistical surveys on a number of topics. Especially efficient for this purposes are the registers of agriculture's enterprises and unincorporated farms which exist in Russia, Belarus, Ukraine, Kazakhstan, Kyrgyzstan and Tadjikistan. Registers of unincorporated farms in Armenia and Moldova have been established recently. Commencing in 2001 the work of establishing these data is to be initiated in Georgia.

The sample surveys which are carried out periodically by statistical offices make it possible to collect data needed for analyses of various aspects of agricultural activities (size and structure of agricultural land, yields of croup products per unit of land, productivity of cattle, availability of fodder, cost of production, availability and use of agricultural machine and others).

In 1993 CIS Statistical Committee sent out to the CIS countries recommendations of the program of agricultural census. This document has been used to some extent by the CIS countries in their activities in this area. During the last 3 years the following censuses were carried out in CIS countries:

- censuses of perennial plantations in Azerbaijan, Belarus, Moldova and Ukraine;
- censuses of cattle in Russia, Ukraine and Moldova;

- census of vineyards in Azerbaijan;
- censuses on various sub-branches of husbandry in Ukraine.

At the present time a preparatory work on agricultural censuses on the basis of program of the World Agricultural Census for 2000 has been initiated. In Kazakhstan, a program of agricultural census to be carried out in 2003 has been tested and a pilot census in one of the districts of Djambul region has been carried out. In Kyrgyzstan, a complete agricultural census is scheduled for 2002.

7. Conclusions

The major conclusions from the above overview are as follows.
Although the CIS countries produce and publish a variety of agricultural statistics, it appears that during the past years of transition from a centrally planned to a market oriented economy no significant progress has been yet achieved in improving these statistics and adapting them to new economic conditions. The quantity and quality of data compiled and published by the CIS countries do not seem to be sufficient to carry out in depth analysis of agricultural economy, to assess many qualitative aspects essential for making sound decisions on important issues of agricultural policy such as the impact of the various types of government support of agriculture (direct and indirect subsidization, loans on favorable conditions, reduced tax rates forgiveness of debt and so forth); overall financial status of agricultural enterprises and their accumulated debt to government, in particular; redistribution of income through price mechanisms, procurement of agricultural goods by state organizations and regulation of procurement prices, the impact of imports of agricultural goods and import duties and so forth.
The emphasis is still put on compilation of indicators in physical units such as production of major agricultural goods, the size of land, yield per unit of land, productivity of cattle and so forth and relatively less attention is paid to indicators of income and finance, cost of production, saving, assets and liabilities, etc. Although, as noted above, all the CIS countries introduced production and generation accounts for agriculture on the basis of the SNA 93, the indicators derived from these accounts (value added, operating surplus/mixed income and so forth) are not sufficiently harmonized with the old indicators that are also published in official statistical editions; the figures on value of output in current price and its disposition on various purposes are computed independently from each other and this creates a problem of their reconciliation.
It is difficult to assess the reliability of the data. There are no longer reasons for the enterprises to overestimate figures on output in the records submitted to the statistical offices (which periodically happened in the USSR), but there is an incentive now to underreport output and profits to avoid paying taxes. The volume of data submitted by the agricultural enterprises to statistical office diminished considerably during the past years of transition; the growing private sector requires the organization of various sample surveys but the resources and experience of statisticians essential for such efficient surveys are not sufficient. As noted above, the agricultural censuses are not carried out as frequently as required and as recommended by the FAO. The economy itself has become more sophisticated and evasive and as a result it is not easy to measure many economic phenomena; though practically all the CIS countries introduce

adjustments for non-observed economy, there seem to be a considerable room for improving the quality of such adjustments.

It appears that insufficient attention has so far been paid to organizational aspects of agricultural statistics; the adoption of special legislation regulating certain important matters, such as the types of surveys and censuses to be carried out and their frequency, the relationship of statistical offices with the respondents, confidentially issues and so forth, would seem to be a very appropriate measure. Such legislation exists, for example, in Germany and studying this German experience in this area would be useful for statistical offices of the CIS countries.

Although there are some positive changes in application of statistical registers and sample surveys, the work in this area is still at the early stages and there is considerable room for further improvements of data collection. In many cases sample surveys are carried out on ad hoc basis and attempts to introduce a systematic approach and planning in this area have so far not been entirely successful.

It appears that insufficient attention to these statistics on the part of the governments or insufficiently clearly defined demand for statistical data from the government agencies responsible for formulation of agricultural policies are the factors which, to some extent, account for the slow progress in transformation of statistics inherited from the USSR times. On the top of that, it appears that agricultural policy has not so far been a priority topic on the agenda of the governments of the CIS countries. Though some countries, such as, for example, Russia, have recently adopted programs of improvement of agricultural statistics, it appears that the scope of these programs and the range of topics dealt with in them are limited and therefore their implementation is unlikely to result in the drastic improvement of these statistics in the near future.

To sum up, the improvement of agricultural statistics of the CIS countries requires a number of interrelated measures pertaining to: i) refining methodology by a more consistent implementation of the international standards, ii) harmonizing indicators characterizing different aspects of agricultural economy with the SNA 93, iii) introducing new indicators essential for qualitative analysis, (analysis of income, productivity, financial aspects, etc.), iv) improving the collection of primary data and especially of the data on activities of unincorporated farms and personal plots of households; this implies a more systematic approach to implementation of surveys and agricultural censuses, v) adopting special legislation which would regulate the most important aspects of organization of agricultural statistics.

Under these conditions it is essential to improve training of personnel, to enhance familiarity of the statisticians with the international standards and experience of other countries. It appears that agricultural statistics should be among the priority topics for the international organizations which provide technical assistance to the statistical offices of the CIS countries.

References

Russian Goskomstat (2000) *Methodology of computation of output, intermediate consumption, gross value added of agriculture in current and constant prices*.

German Federal statistical office (2001) *New stage of development of agricultural statistics*

Challenges of the Agricultural Statistical System in Hungary

Éva Laczka
Hungary Central Statistical Office, Agricultural Statistics Department,
H-1024, Budapest, Hungary, Keleti Károly u. 5-7,
e-mail: eva.laczka@office.ksh.hu

Abstract: Agriculture played an important role in our country's economy in the last century and the same can be said today. The Hungarian statistical service has been collecting agricultural data since 1867. Both the statistical service of the last century and statistical service of the present century observed the data of economic organisations which was said to be the most important. In recent years fundamental changes have taken place as regards the ownership and the structure of agriculture in Hungary. The elaboration i.e. development of our statistical data collection system required by the changes in the range of data suppliers had to be realised along with the adaptation of international standards. Working on it, the major challenges are: demands of the domestic users; liaison with the users; increasing demand on information and the respondent burden; methodology development such as the investigation of applicability of new types of data collection and new international methodology standards.

Keywords: law on statistics, international standards, domestic users, division of labour, respondent burden, methodology development.

1. Hungarian Agriculture

Due to its natural conditions and historical traditions agriculture and food industry play a significant role in the economy of Hungary. The cultivated agricultural land area is 70% of the total land area, while 8% of the active earners are employed in agriculture. The share of agriculture in the Gross Domestic Products (GDP) is about 5%, and that of the food industry is more than 6% (Figure 1).
In recent years fundamental changes have taken place as regards the ownership and the structure of agriculture in Hungary. In previous years there were a relatively small number of large scale agricultural and food-industrial holding. The land area, state farms and the transition of agricultural co-operatives to a large number of small and medium size agricultural units were established on which far less data are available. The elaboration i.e. redevelopment of our statistical data collection system required by the changes in the range of data suppliers had to be realised along with the adaptation to international standards.

Figure 1: *Macroeconomic Share of the Agricultural Sector in Hungary*

Year	Percentage share of agriculture in					Export-import balance a/
	GDP Production	Consumption	Exports a/	Investments	Employment b/	
	At current prices (%)					
1994	6,0	34,0	20,6	2,6	8,7	111,9
1995	5,9	34,5	22,0	3,0	8,0	227,6
1996	5,8	33,1	18,4	3,5	8,3	244,7
1997	5,2	33,3	13,0	3,6	7,9	295,6
1998	4,9	33,1	10,5	3,6	7,5	314,7
1999	4,2	32,5	8,0	3,3	7,1	273,3
2000	..	..	6,9	2,9	6,5	302,2

a/ Agricultural and food products.
b/ Labour survey data.

2. The legal background of data collection

Six Laws on statistics including provisions for agricultural statistics have been passed in Hungary so far. The first Law on statistics whereby the Parliament authorised the annual data collection plan was enacted in 1873.
After major amendments during the following decades several Laws on statistics followed, such as the Law on statistics of 1973 which provided a standard framework for the statistical activity. The two components of statistical activity defined in the Law were the administrative (state) statistics and official statistics. The entities responsible for data collection for the latter were the ministries and agencies of nation-wide authority. In this sharing of efforts HCSO was responsible for co-ordination and professional oversight. The objective was to avoid duplication of statistical effort.
The latest Law on Statistics was enacted in 1993 and amended last in 1999. Instead of the administrative (state) and official statistics it stipulates the so called statistical service with members including the HCSO, the ministries and some agencies of nation-wide authority. First the National Statistical Council deliberates the draft of the annual data collection plan prepared by the statistical service. Apart from the constituents of the national statistical service unions and representatives of science delegate members to the National Statistical Council. Upon their acceptance the president of HCSO submits the annual statistical data collection plan for government approval.
In the Law on Statistics now effective, major emphasis is given to privacy (data protection) issues and the disclosure of data of public interest. The staff of HCSO and other members of the statistical service currently work on the data collection plan for year 2002.
Some of the major challenges facing the authors of the data collection plan for agricultural statistics include

International standards;
Specific demands of the domestic users;
Sharing of effort between HCSO and the competent authority;

Increasing demand on information and the respondent burden;
Liaison with the users, and
Methodology development such as the investigation of applicability of new types of data collection and new international methodology standards.

3. Challenges in the field of agricultural statistics

3.1. International standards and expectations

Excluding the period between 1953 and 1968, Hungary has been a member of **FAO,** consequently the FAO methodology standards governed the design of agricultural census plans for many decades, including the preparatory efforts to the agricultural censuses of years 1972, 1981, 1991 and 2000.
In addition to censuses Hungary submitted 10 reports to FAO each year.
A strategy up to year 2000 was drawn up following the definition in 1994 of a framework for co-operation between **Eurostat** and HCSO, with a purpose to meet the harmonisation obligations. As part of the accession negotiations, the screening of the Chapter on Statistics took place in July 1998. The Hungarian party declared that Hungary would adopt the *acquis* in the field of statistics without any need for derogation.
During the preparatory efforts the Department of Agricultural Statistics of HCSO analysed the level of harmonisation of the Hungarian system of agricultural statistics in 1999 (Figure 2).
In agreement with Eurostat the agricultural census of year 2000, the census of vineyards and orchards of year 2001 and the design of the system of Agricultural Accounts were selected as the key tasks. In addition to these comprehensive efforts HCSO reports data to Eurostat on 30 occasions yearly.
Hungary has been member of the **OECD** since 1996. Adoption of the methodology standards and systematic reporting also form part of the membership obligations.
In the past few years the above mentioned international organisations made substantial efforts in the approximation of methodology standards and reduction of duplicate reporting, specifically in the field of recommendations concerning censuses.

Figure 2: *Division of labour between HCSO and MARD, to harmonize of agricultural statistics according to the chapters of the compendium (March 2001)*

Form of regulation Level of harmonisation	Juridical obligations	Gentlemen's agreements	A purely voluntary base
Mainly harmonised	621. Farm structure survey 621. Community typology for agricultural holding 623. Statistics on viticulture 624. Statistics on fruit growing 641. Corp production statistics 651. Livestock statistics 681. Forestry statistics 694. Aquaculture statistics	611. Land use statistics 633. Agricultural price statistics 634. Agricultural labour input (ALI) statistics	
Partly harmonised	612. Remote sensing 652. Animal products statistics FADN	631. Economic accounts for agriculture 642. Supply balance sheets for crop products 643. Harvest forecast (agromet) 653. Supply balance sheets for animal products	
New objective	622. Eurofarm database	632. Income of the agricultural households sector 654. Fodder balance sheets	635. SPEL 661. Good industry

Perhaps the system of agricultural accounts is the field where international methodology standards are least harmonised. Eurostat recently adopted the new EAA 97 methodology. The OECD methodology was practically identical with the old Eurostat methodology, but OECD has not adopted the new one yet. The standards in FAO's methodology manual – the SEAFA – also differ from those of the previous two international organisations. In our view the relatively minor differences in methodology allowed for standardisation which, in turn, would alleviate the burden of statisticians of the member countries.

3.2. Peculiarities and Hungarian user demands

Perhaps agricultural statistics has the richest history in the 130 years chronicle of the independent Hungarian statistical service. Time series of over 100 years are available on the use of land and animal breeding. In the 50's, 60's and 70's of the 20th century most of the agricultural output was produced by state farms and collective farms, thus comprehensive and systematic surveys covered these economic units. Due to their relatively small share, the agricultural activity of private holdings was mainly estimated on the basis of the administrative sources. (Figure 3). Since the share of animal breeding started to increase in the private farms from the end of the 50's, HCSO implemented in 1956 a representative census on animal breeding of households.

Up to 1995 HCSO conducted quarterly sample surveys on cattle, pigs, sheep and poultry.

From 1996 sample surveys on animal breeding were conducted every four months for specific species in accordance with the relevant EU standards. By the time of the agricultural census of year 2000 it became clear that data collection at reduced frequency not only caused discontinuity in the time series but also failed to meet the needs of Hungarian users. It was specifically troublesome e.g. in pig breeding, but in other areas as well.

For this reason after the census of year 2000 **the cattle, pig, sheep and poultry population has been surveyed every four months on the basis of a representative sample.** Nowadays the need for quarterly representative animal breeding surveys has cropped up, indeed in compliance with the relevant EU regulations and criteria. The data input for quarterly GDP calculations is also a requirement in favour of the quarterly animal surveys.

3.3. Sharing of efforts in the field of agricultural statistics

The key user and one of the most significant producers of agricultural statistics – apart from HCSO – is the Ministry of Agriculture and Rural Development (MARD). In 1999 the leaders of the two organisations signed an agreement on sharing the efforts (Figure 2), which has been effective since that date.

According to the present Law on statistics within the frame of the official statistical service, the HCSO has the main responsibility concerning agricultural statistics. Mainly operative informations are collected by the MARD. Fields of the activity performed by MARD are: estimations, data collection concerning the cost of agricultural production. In addition MARD is also responsible for forestry, fishery statistics and remote sensing.

3.4. Increasing demand on information – respondent burden

In Hungary 8500 economic organisations and 960 thousand holdings are engaged in some kind of agricultural activity, therefore they are respondents of the system of agricultural statistics. Economic organisations are obliged to mail reports several times each year. 70 thousand of the 960 thousand households are included in the sample surveys (Figures 3 and 4). This burden is specifically heavy at the turn of the millennium, when enumerators visit respondents on account of the agricultural census, then for the population census, and then for the vineyard and orchard census of year 2001.

Due to the increasing demand on information no group of questions can be deleted. Individual data collection exercises cannot be replaced by data borrowed from administrative sources due to their current unsatisfactory quality. The only alternative is the use of well-designed data collection systems and questionnaires.
This was the philosophy behind the **„integrated December survey"** developed by HCSO this year, which amalgamates the sets of indices of three earlier representative surveys in a reduced volume, and not at the expense of the quality of the input information.
Another important consideration is to "engage" the respondents. For this purpose HCSO produced the first "calendar" for those involved in the yearly systematic supply of data. Apart from the exoteric explanation of the scope and aims of statistical censuses and surveys, this brief publication also provides useful agricultural information.

3.5. Liaising with the users

Apart from the user contacts stipulated in the Law on statistics, the agricultural census of year 2000 opened ways to build contacts of new type.
Due to the radical changes that took place in the 90's of the 20^{th} century eager expectations preceded the results of the census of year 2000. The "Reckoning Club" evolved from discussions following press conferences and other media events. The participants are statisticians, agricultural experts, members of the product councils and agricultural chambers, and representatives of the media. The "Reckoning Club" holds ad-hoc meetings for the discussion of specific topics. In the quarterly newsletter of the Club articles are published by the representatives of the agricultural profession.

3.6. Methodology development

The agricultural census of year 2000 and the vineyard and orchard census of year 2001 are the two key tasks of the new ones to be carried out upon the recommendations and standards of international organisations.
The vineyard and orchard census of year 2001 is of paramount importance because no similar census has been conducted in Hungary since 40 years. The census is expected to cover 800 economic organisations and 300 thousand households.
Though monitoring of agricultural services is a task of lesser size, it is still rather complex because of the novelty of this activity, which only appeared a few years ago in the Hungarian agriculture.

Figure 3: *Agricultural Statistical System 1950-2000*

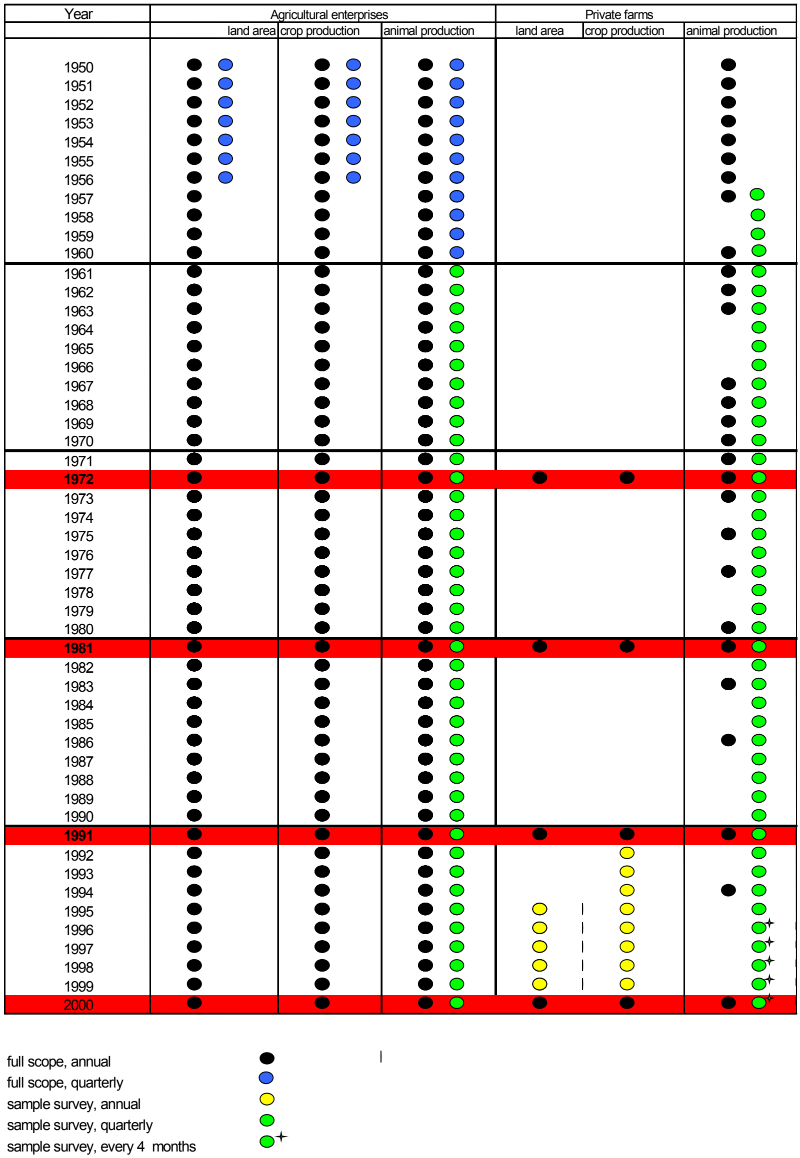


Figure 4: *Data collection system 2001. (Agricultural production)*

Denomination	Data collection regarding to		Total coverage, %	Expert estimation
	enterprises	private farms		
Land area and sown area, 31 May	annual	annual	about 85	
Yield of cereals	annual	annual	about 90	
Yield of main crops	annual		about 85	
Livestock (1 April, 1 August, 1 December)	every four months	every four months	about 99	
Poultry hatching	monthly	monthly	100	
Milk statistics[a)]	monthly	monthly	about 95	

December survey

Production of crops and vegetables, utilisation and net sales	annual	annual	about 85	
Fruit and vine-growing, -planting, -grubbing, -utilisation and net sales	annual	annual	about 85	
Changes in livestock, production of slaughter animals and animal products, utilisation and net sales	annual	annual	about 85	

Full scope — Sample survey (sample size 6%) — Expert estimation

a) enterprises engaged in milk production or processing activity

Among the new tasks the system of Agricultural Accounts is worth mentioning. HCSO has been developing this system since 1999, when the Section of Agricultural Accounts was established at the Department of Agricultural Statistics. It is expected that all data in the entire sets of accounts for the past few years will be available by the fall of 2001 (Figure 5).

Figure 5: *Data collection system 2001. (Economic Accounts for Agriculture)*

Denomination	Data collection regarding to		Total coverage, %	Expert estimation, %
	enterprises	private farms		
Expenditures of agricultural producers	annual	annual	80	
Report on procurements[b]	annual		about 70	
Livestock market report		monthly (weekly)	about 40	
Consumer market report		monthly (weekly)	about 40	
Balance of agricultural trade[b]	annual		about 80	
Sales of fertilizers[c]	annual		about 85	
Sales of pesticides[c]	annual		about 85	
Volume and value of feedingstuff sales[c]	annual		about 85	
Sales prices of veterinary products[c]	annual		about 85	
Sales of industrial products used for agricultural production[c]	annual		about 85	
Sales prices of feedingstuffs[c]	quarterly		about 60	
Sales prices of fertilizers[c]	quarterly		about 60	
Sales prices of pesticides[c]	quarterly		about 60	
Sales prices of veterinary products[c]	quarterly		about 60	
Agricultural services	annual	annual	about 95	
Agricultural Labour Input		annual	about 60	

Full scope — Sample survey (sample size 6%) — Expert estimation

b) Wholesalers and processing organisations
c) Retailers of means of agricultural production

Due to the characteristics of the Hungarian structure the adoption of some EU methodology standards poses a major challenge for the community of Hungarian statisticians.
In the case of the agricultural enterprises there is no ambiguity concerning the need for annual surveys, classification and/or inclusion of farms in the EAA or the EU analyses.
Based on the 2000 census data, however, we must re-consider some of the issues concerning households and small farms.
We must carry out calculations based on the 2000 Agricultural Census in order to define the size of Hungarian farms where production solely covers own consumption. The objective is to arrive at the size categories where either surplus is produced or where the prime objective of production is the sale of products, that is, to determine the market producer farms.

References

Law on statistics No XLVI of 1993, Hungarian Central Statistical Office, Budapest 1999.
Statistical Yearbook of Agriculture, Hungarian Central Statistical Office, Budapest 2000.
Hungarian Statistical Review, vol. 4. 2000, Budapest 2000.
The Agricultural Census 2000, (CD-ROM) Hungarian Central Statistical Office, Budapest 2000.
Hungarian Agriculture 1851-2000, (CD-ROM), Hungarian Central Statistical Office, Budapest 2000.

IWG.AGRI: Past, Present and Future

Calò, Giuseppe
Eurostat, Jean Monnet Building - BECH B2/472, Rue Alcide de Gasperi, L-2920
Luxembourg - Kirchberg, e-mail: giuseppe.calo@cec.eu.int

Karlsson, Jan
UN/ECE, Palais des Nations - 8-14 Avenue de la Paix, CH-1211 Geneve 10,
Switzerland e-mail: jan.karlsson@unece.org

Lindner, Andreas

OECD, 2, rue André Pascal, F-75775 Paris Cedex, 16, Paris, France
e-mail: Andreas.Lindner@oecd.org

Muthmann, Reiner
Eurostat, Jean Monnet Building - BECH C3/631, Rue Alcide de Gasperi L-2920
Luxembourg - Kirchberg, e-mail: rainer.muthmann@cec.eu.int

Narain, Pratap
FAO, Via delle Terme di Caracalla, 00153 Rome, Italy
e-mail: pratap.narain@fao.org

Abstract: The Inter-Secretariat Working Group on Agricultural Statistics[1] (IWG.AGRI) was created in 1991 to help transition economies to adapt their agricultural statistical systems to the needs of a market economy. Now, a decade later, at the dawn of a new millennium, it seems appropriate to take stock of past achievements, present issues and identify possible future action of the IWG.AGRI.
Key issues in this paper are the seven seminars organised by the IWG.AGRI, the impact of the enlargement process of the European Union and the consequences of joint data collection and data sharing in the electronic world.

Keywords: IWG.AGRI, transition economies, agricultural statistics, joint data collection, international co-operation and co-ordination, technical assistance programme, Central and Eastern European Countries.

---

[1] The IWG.AGRI was established in 1991 by the Organisation for Economic Co-operation and Development (OECD), the European Union (EU-EUROSTAT), the United Nations Economic Commission for Europe (UN-ECE) and the Food and Agriculture Organisation of the United Nations (FAO).

1. Past

1.1. General

After the political and economic changes in the early 90s in Eastern Europe and in countries of the former Soviet Union, the main concern of international bodies, operating in the field of agricultural statistics was to find an appropriate way to adapt efficiently to this new situation. It was clear from the outset that a well-structured, non-duplicative reinforcement of their co-operation was the only possible way to provide a tailor-made support programme to these economies in transition to smoothen their integration into existing statistical systems. This "reinforcement" resulted in the "Inter-secretariat Working Group on Agricultural Statistics" (IWG.AGRI) comprising representatives of the UN/ECE, Eurostat, FAO and OECD.

The philosophy behind the IWG.AGRI was, and still is, to co-ordinate the technical assistance programmes in the domain of agricultural statistics for Central and Eastern European Countries (CEEC) and New Independent States (NIS), to avoid overlap or duplication of activities, and to progressively integrate these countries into the analytical work of the respective organisations. One of the key instruments to achieve this goal was to organise workshops or seminars[2] on agricultural statistics. On a rotation basis, each of the four IWG.AGRI members took the responsibility of the organisation, initially only meant for CEEC and NIS countries. A particular feature of these seminars was that the IWG.AGRI organisations not only shared costs, but also complemented each other with respect to geographical coverage of attendance, financial support, including support in kind (e.g. free translation of papers into Russian).

1.2. The seven seminars

The first four seminars had to be of a more general nature, covering a broad range of issues. The main goal was to exchange basic information and expertise covering agricultural statistics and information systems in order to get a comprehensive picture of the most problematic areas for immediate and future action. It should be underlined that the agenda list of topics was elaborated by the IWG.AGRI in close consultation and co-operation with the host country.

In recognition of the progress made so far in transition economies, the 1997 Budapest Seminar for countries in Central and Eastern Europe and also the 1998 St Petersburg Seminar for CIS marked the end of IWG.AGRI Seminars covering more general and basic topics (= horizontal focus). In Budapest, two major areas of agricultural statistics were chosen for an in-depth discussion: EU's Agricultural Census 1999/2000 and Agro-Monetary Statistics (AMS) with particular emphasis on the Economic Accounts for Agriculture (EAA) as a satellite Accounting system linked to the SNA93. Budapest marked the beginning of a vertical, subject matter driven, focus.

The St. Petersburg Seminar was the first seminar with a clear focus on the regional dimension. The target group was the Russian Federation and its Regions in their efforts to adapt the national agricultural statistical system to the changed needs of a market economy. The range and diverse nature of problems identified required this approach. Key issues were: how to adapt agricultural statistics to structural changes, the development of pertinent indicators, agricultural accounting, the design of structural

[2] Sofia (1992), Warsaw (1993), Minsk (1994), Brdo (1995), Budapest (1997), St. Petersburg (1998), Luxembourg (2000), Paris (2002).

surveys and sample surveys, general methodological problems and the role of information technology.
St. Petersburg marked the end of IWG.AGRI Seminars held in transition countries. In recognition of the -sometimes- considerable progress made in transition countries to adapt to the new paradigms and in pursuing the objective of keeping costs to the strictest minimum, IWG.AGRI - members decided to organise henceforth the forthcoming seminars

a) back to back with existing statistical meetings at the location of an IWG.AGRI member, and

b) to focus on a policy/analytical relevant theme by "drilling down "vertically instead of covering horizontally all major subjects.

The first of these new Seminars took place in Luxembourg, 5-7 July 2000. It focussed on Agricultural Economic Statistics[3] (and was christened AgES) with an in-depth review of: Economic Accounts for Agriculture, Agricultural Sector Modelling, Agricultural Price Statistics, Agricultural Income Indicators and Agricultural Labour Input.
Pending confirmation, an 8th IWG.AGRI might be held at OECD in Paris in 2002 focusing on agricultural statistics for policy makers.

2. Present

2.1. The audience of the seminars

As already touched upon in the last paragraph, due to different objectives, location, specific problems to be addressed, the audience of the seminars changed over time. In St. Petersburg, on invitation of the OECD, a Chinese delegation participated. For several reasons in Luxembourg, the AgES seminar attracted delegates from forty different nationalities world-wide.
Firstly, it was organised back to back with an EU Working Party meeting on EAA (same topics as the seminar) which made it attractive for EU and CEEC delegates to attend the seminar as well.
Secondly, Eurostat currently supports a huge Pilot Project for Phare CC's on Agricultural Monetary Statistics (AMS) to bring them in line with the EU-standards. The topics of the 7th Seminar formed a perfect match with this project and attracted additional participants.
Thirdly, bearing this in mind, it followed to invite countries with transition economies from outside the traditional target group and optimise the benefit from the IWG.AGRI experience in this domain.
Finally, Agricultural Economic Statistics form an important basis for (policy) analysis. OECD's revision work on the EAA, FAO's world-wide handbook SEAFA and recently the adoption by the EU of the revised EAA methodology (based on the principles set out in the SNA93 and the ESA95), opened the way towards better and comparable

[3] AgES was the name of this 7th seminar being the acronym of Agricultural Economic Statistics.

(agricultural) information systems on a global level. Because of this "chaining of methodologies", the AgES programme also became attractive for more advanced nations.

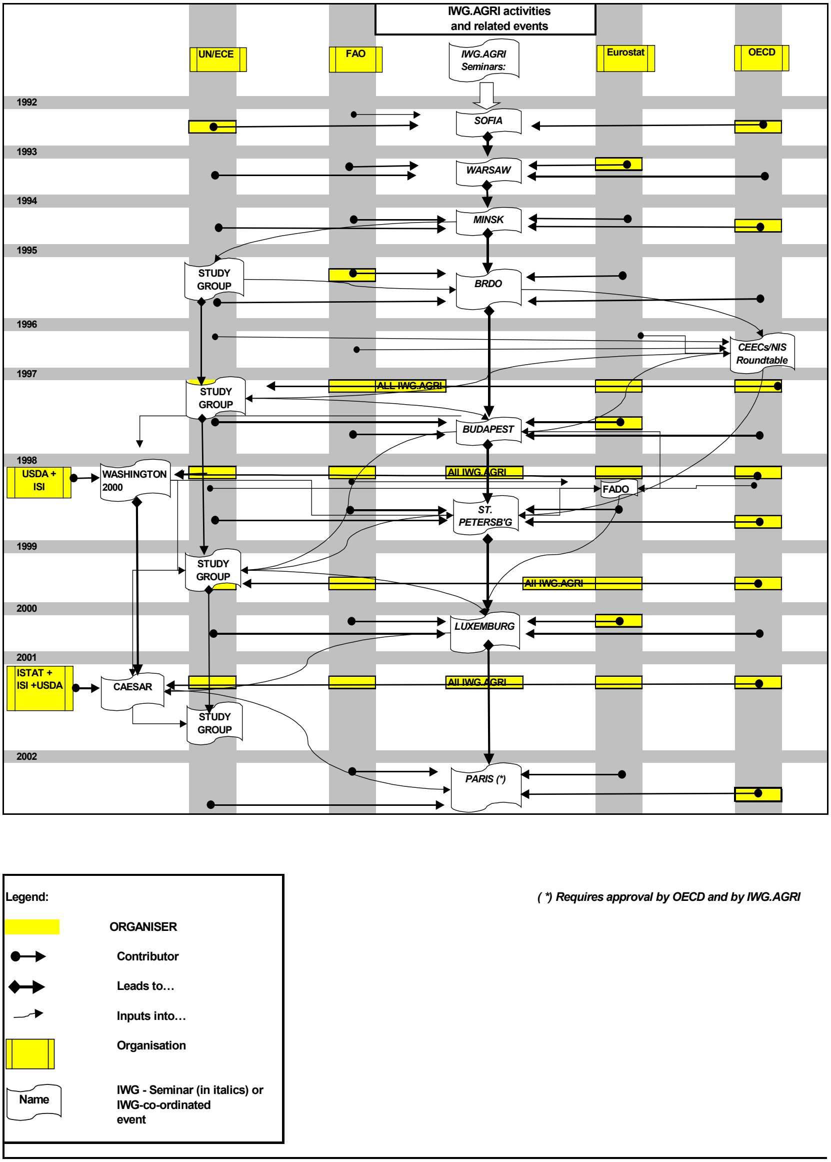
2.2. Other activities of the IWG-AGRI

Figure 1 (see below), which has been presented to the UN/ECE meeting in Geneva in 1999 and to the CES in 2000, gives a comprehensive summary of all the IWG-AGRI meeting activities. Its particular merit is that it illustrates the existing linkages between apparently separate or isolated meetings. Studying this chart makes it clear that the IWG.AGRI is literally the driving force behind all major international events since 1992. It was one of the major players behind the organization of the first World Conference on Agriculture Statistics (Washington DC, 1998). It also took the initiative of creating an organizing committee for the 2nd World Conference - Conference on Agricultural and Environmental Statistical Application in Rome, CAESAR which is hosted by ISTAT and takes place in Rome on 5-7 June 2001. Back-to-back with CAESAR, satellite meetings for transition countries (4 June, organised by IWG.AGRI) and for developing countries (8 June) are foreseen.
IWG.AGRI, especially Eurostat, FAO and OECD, are also providing assistance to China (and in this field Italy is very active), in particular in conjunction with the agricultural census and EAA. With almost 20% of GDP and over half the population in rural areas/agriculture, the dimension of the problems is such that a sustained and co-ordinated co-operation between IWG-AGRI and China is necessary.

2.3. Joint data collection

Scarcity of resources and budgetary constraints have given new impetus to exploring feasible and satisfactory ways of lowering the statistical response burden for National Statistical Bodies vis-à-vis International Organisations. In addition, there have been growing pressures to identify and clarify common areas of regular statistical needs across International Organisations, not only to avoid unnecessary duplication in requests to National Statistical Bodies, but also to reduce discrepancies in published data and to explore scope for joint activities.
A number of joint data collections already exist amongst International Organisations in agriculture. For instance, OECD and Eurostat closely co-operate on the EAA without duplication.

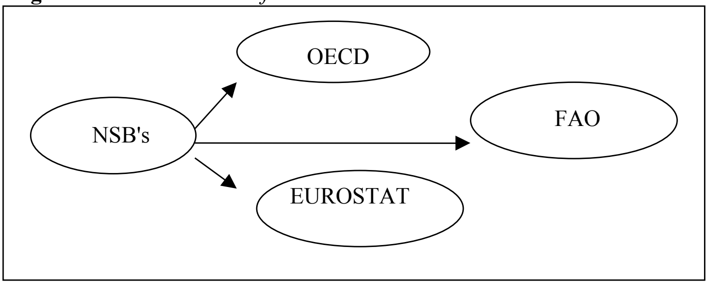
Figure 1: *International co-ordination of agricultural statistics*



*(*) Requires approval by OECD and by IWG.AGRI*

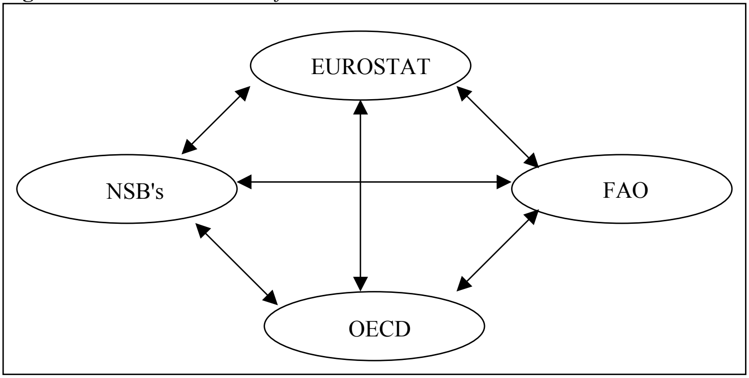
When discussing joint data collection, great care has to be taken to distinguish between regular and new and/or ad-hoc needs (which may become regular needs later on), and between International Organisations requirements for Member countries and for Non-Member countries (in the field of agriculture they are not necessarily the same), as well as the different time-scales of needs across International Organisations. Therefore, any sharing agreement has to be looked upon on a case by case basis.
IWG.AGRI has developed a detailed model for how joint data collection/data sharing can be achieved (see figures 2 and 3). Although not yet implemented, it was discussed by member countries during the 1999 ECE/Eurostat/FAO/OECD meeting in Geneva and approved by participants.

Figure 2: *The INPUT data flow*



Parallel to this, IWG.AGRI has also carried out an in-depth survey among a few member countries in order to explore the volume and content of the requests of data from International Organisations and the degree of duplication although neither the response burden nor duplication was felt to be an immediate problem.
As for the EU candidate countries a large amount of requests originated not only from OECD, FAO and Eurostat but also from "Brussels". With the candidacy also followed the requests for long time series data which was felt being in particular burdensome as they also cover the early period of the transition period when data were scarce or incomplete. In this context it is also interesting to note that even if (International) Statistical Services try to harmonize data collection there are many surveys that they have very little control over. Institutes like Ministries, Tax Authorities and National Banks, posses and maintain information systems with a specific purpose but a (inter) national character.

Figure 3: *The OUTPUT data flow*



Another activity initiated by IWG.AGRI with the view of facilitating joint data collection is the Handbook of Concepts and Definitions Used in International Collections of Food and Agriculture Statistics.
Similarly, IWG.AGRI is currently compiling all available proceedings of IWG.AGRI seminars on CD-ROM and an Internet Web-site for dissemination and consultation. This initiative will provide a unique statistical knowledge base for both producers and users of agricultural statistics.

3. Future

Although much has been accomplished as concerns assistance to countries in transition, many of which have significantly raised the quality of their agricultural statistics in the last few years, there are still a number of countries, the NIS in particular, that need further "tailor-made" assistance of different kind.
At the same time, the IWG.AGRI members continuously strengthened their inter agency co-operation. This increasingly became an indispensable element in their work for two major factors:
First, the transition process induced a dynamic by which "territorial identities" of international bodies changed as well. FAO and UN/ECE received new, independent countries, OECD now has no less than four new Member countries which before were "behind the iron curtain", and the EU is engaged in an enlargement process of a formidable dimension and consequences.
Second, the exponential development of IT applications and, above all, the Internet revolution offer completely new and innovative ways of collecting, processing and disseminating data.
Having said that, the conclusions of the most recent joint meeting on Food and Agriculture statistics as well as those of the AgES Seminar, lead to the following list of activities of the IWG.AGRI in the years ahead of us:

(1) The Agricultural Census (world-wide) and the Farm Structure Survey

(2) More work in the area of standardisation, similar to the Handbook, which was mentioned above.

(3) Focus the work on the areas of agro-business/agricultural economic statistics;

- interaction of agriculture with environment and rural development;
- use of administrative and other non-survey data;
- more in depth studies of income statistics;
- productivity measurements; and
- measurement and forecasts of food supplies.

(4) The use of new technologies for collecting, processing and distribution agriculture statistics.

(5) With the help of new technologies to further pursue joint data collection and joint data consultation.

(6) Development of meta data and methodological sheets for future harmonised development of derived statistics

(7) A framework for development of spatial data and decision support tools including Geo-Networking

It is understood that parallel to the above joint activities of ECE, Eurostat, FAO and OECD, each organization has a programme of work, which is focusing on the specific needs of the member countries of each respective organization.

There is also a need to strengthen the analytical part of the presentation of statistical data. More investigation is needed, although this has not yet started, on the reasons for differences between published data. Too often, a consumer is confronted with different data for the same variable and year, depending on which publication and organizations is consulted. Here, technology might facilitate things. Already now, the Web pages of the organizations of IWG.AGRI are linked together. In the future, IWG.GRI will try to alert differences in data and give explanations why there are differences.

IWG.AGRI pays a lot of attention to the use of technology not only from the latter point of view but also in the whole process from the collection of data in member countries to their processing and distribution and integration. There is no doubt that technology will play a much larger role in the production of the statistics in the next few years.

Information, which was until recently a monopoly of public Statistical Services, is nowadays in principal accessible (Internet) by everybody in the world. Many users see this as an important step to transparency and freedom of information but paradoxically this is also the most alarming threat. A growing lack of background information by users (e.g. meta-data, concepts and definitions) can and may lead to the uncontrolled growth and abuse of the data with the opposite effect: in-transparency of and confusion around the results.

It follows that the IWG.AGRI should redefine both its role in and commitment to this international statistical process and perhaps seek a more active role in this respect, in particular in the area of standardization.

The process of integrated questionnaires and exchange of information have a high priority on the agenda's of the International Organisations and already some time ago the IWG.AGRI presented concrete proposals around how joint data collection can be managed via the Internet[4].

In a paper, recently presented by the FAO[5] some very interesting ideas and suggestions around joint data collection, data sharing and data distribution can be found. An excellent example is given: forestry statistics. Here considerable progress has been made towards common questionnaires. The Forestry Department of the FAO is currently executing for the second time a Joint Forest Sector Questionnaire with Eurostat, ECE and the International Tropical Timber Organization (ITTO). This builds upon more than ten years of a joint forest product questionnaire executed by FAO, ECE and Eurostat. However, each of the organizations publishes the data independently according to their specific needs.

FAO and Eurostat recently started investigating problems and deviations of common but conflicting data in the agricultural domain, another direct result of the need of streamlining data-flows. Here, more than data quality the (reduction of) statistical burden (as mentioned before) brought the ideas behind joint data collection into practice.

Mr Ivanov's paper[6] reviews the demand for (statistical) assistance by NIS countries. Going through his shopping list it is clear that, apart from joint data collection, probably IWG.AGRI's most important role is training in the broadest sense of its meaning. Furthermore, it is interesting to see that the topics on his list in fact cover the demand of many countries in the world:

"The improvement of agricultural statistics of the CIS countries requires a number of interrelated measures pertaining to:

- *Refining methodology by a more consistent implementation of the international standards.*
- *Harmonising indicators characterising different aspects of agricultural economy with the SNA 93.*
- *Introducing new indicators essential for qualitative analysis, (analysis of income, productivity, financial aspects, etc.).*
- *Improving collection of primary data, especially on activities of unincorporated farms and personal plots of households. This implies a more systematic approach to implementation of surveys and agricultural censuses.*
- *Adopting special legislation, which would regulate the most important aspects of the organisation of agricultural statistics."*

[4] Geneva, 23-25 June 1999, CES/AC.61/1999/19 Meeting on Food and Agricultural Statistics in Europe. "Joint data collection/data sharing between International Organisations.

[5] Washington D.C., 20-22 September 2000, ACC Subcommittee on Statistical Activities. "Quality assurance, arrangements with common questionnaires and conflicting statistics".

[6] Rome, 4-8 June 2001, Agricultural statistics in the CIS countries, general overview. Mr. Youri Ivanov, Deputy chairman CIS STAT Moscow, Russian Federation.

One of his other conclusions is that *"It appears that agricultural statistics should be among the priority topics for the international organisations, which provide technical assistance to the statistical offices of the NIS countries."* An important tool *"To enhance familiarity of the statisticians with the international standards and experience of other countries"* underlines the coaching role of the IWG.AGRI.

4. Conclusion

The history of the IWG.AGRI can be characterised as dynamic, user-oriented and successful. In a short period, covering not more than ten years, this co-operation has proven to be solid, efficient, creative and flexible not in the least due to the commitment and professional approach of individuals engaged in this joint effort. .

The next decade however, probably requires an even higher commitment. Although a part of Central and Eastern European Countries moves into the direction of the EU umbrella, an increasing demand for support from other regions in the world will easily compensate this "loss".

With the introduction and the success of the Internet, the risk of loosing the monopoly (and control) of statistical information systems is evident and partly unavoidable. National and International (statistical) Organisations have the obligation to guarantee the quality of this data.

A prerequisite to manage and guide these processes in the agricultural statistical domain is a solid and sustainable IWG.AGRI. A strong mandate by all partners, with the support of a (small) permanent professional staff, makes is possible to anticipate on the rapid changes and handle the tremendous variety of demands to at least survive another ten years.

Istat Activities in Candidate and Transition Countries

Salvatore Favazza, Micaela Jouvenal
ISTAT – International Relations, Via A. De Pretis, 77, 00187 Roma, Italy
e-mail: favazza@istat.it, jouvenal@istat.it

1. Background

ISTAT's international activity is progressively and constantly growing and plans its participation in the international statistical community according to both the new opportunities offered by the evolution of the reference framework, and the growing in-house experience and consciousness of the importance to be part of the international processes, not just as a simple spectator, but with an active role.
Certainly, the intensity and quality of the required engagement is strictly related to the participation of Istat to the decision processes for the construction of the European Statistical System; in fact, the aim of harmonising and integrating statistical information to properly support the European Union's policies, also requires a systematic involvement of Istat's staff into international activities together with the colleagues of other Statistical Institutes.
Furthermore, such a type of involvement is not limited to the European Union context, but is extended to other international organisations such as OECD, the UN Statistical Commission and UNECE, IMF, World Bank, FAO, etc., whose Working Groups are more and more requesting statistical competence as a key element to plan and evaluate institutional activities.
Within this framework, technical co-operation in the field of statistics is playing a role of increasing importance in Istat, for the implementation of Italian bilateral programmes, as well as for those programmes financed by international organisations, and particularly by the European Commission. This sector, which has enlarged the geographical reference context over the last ten years, absorbing an increasing number of qualified resources specifically selected for their competence and technical capabilities, has been instrumental to acquire sensitiveness and skills to operate within an international context.

2. Istat's technical co-operation strategies and orientations

Since 1992 Istat has decided to participate in the international statistical co-operation processes, launching a programme of activities that was strictly related to the Italian foreign policy, establishing links with the Ministry of Foreign Affairs – Directorate General for Co-operation to Development, and to the European co-operation programmes, by participating in Eurostat's Working Groups and in co-operation projects financed by the Commission.
Having chosen these as the main reference parameters for its co-operation strategies, activities have also been implemented with the financial support of other international organisations such as the World Bank, the IMF and several other UN agencies (FAO, UNICEF, etc.).

Istat's implementation strategies, initially aimed at responding to countries' specific requests, later followed an ongoing definition process based on two major aspects: geographical areas of intervention and technical subject matter areas. Today, geographical priorities are oriented to support the statistical systems of the countries of the Mediterranean basin and of Balkan and Accession countries; attention is also given to the Russian Federation, China, and the MERCOSUR countries. Sectorwise, Istat is concentrating its efforts mainly in the field of demographic and social statistics, in particular to the analysis of poverty, the estimation of the Non-Observed Economy and Agriculture statistics. To support and strengthen the institutional capabilities of the recipient institutions, assistance is also provided for the definition of dissemination policies, for the drafting of statistical laws, and for the computerisation of statistical processes.
Istat's involvement has also in time being influenced by the evolution of policy approaches followed by its funding organisations: on the one hand, the Italian Ministry of Foreign Affairs has supported the implementation of programmes dedicated to the institutional strengthening of the statistical systems of selected countries of interest, regardless of the specific area of intervention, the approach of the European Union has experienced different phases over the last years, starting from a purely demand driven approach, which partly remains the case for the transition countries of the New Independent States and Mongolia, to the accession driven approach presently followed for Candidate Countries, with programmes aimed at reaching statistical compliance to the requirements of the *acquis communautaire*. Somewhere in between lie the Balkan countries, which have largely benefited from the Phare programme, today exclusively dedicated to Candidate Countries. Although Balkan countries will soon fall under the newly established CARDS programme, which will privilege other aspects – like, for example, the regional approach – these countries, and especially Albania and Macedonia, have followed for many years the activities planned within Phare, including participation to several Eurostat's working groups and task forces; however today, these changed policy approaches as well as the focus on Candidate Countries have basically largely decreased the interest of member States to work with non-candidate countries. In contrast with this, Italy remains strongly involved in this neighbouring area, thus allowing Istat to implement relatively large bilateral projects there.

3. An overview of technical assistance to transition countries

The Table in the Annex provides a complete list of the activities implemented by Istat over the past eight years to assist the statistical systems of the Transition countries, specified by partner country, by area of intervention and by number of days and Istat staff involved. The following summary Table 1, indicates that approximately 1,600 man/days of technical assistance have been spent, either in the partner country or at Istat during visits of delegations or at the occasion of co-ordination meetings; this results in a yearly engagement of approximately 200 man/days, which for an Institution like Istat, "relatively" new to co-operation issues, represents a strong engagement.

Table 1: *Summary of Istat assistance to transition countries, 1993 – 2001 (June)*

Countries	Fields of activitiy	N° of days	Istat staff	man/ days
18	24	1.002	183	1.595
Albania, Armenia, BiH, Czech Republic, Fyrom, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lituania, Moldova, Poland, Romania, Russian Federation, Slovak Republic, Ukraine, Uzbekistan	Labour Force, Agriculture, Demography, Cartography, Censuses, Non Observed Economy (NOE), Prices, Households Budget Survey, National Accounts, Classifications, SME, Statistical methodologies, Trade, Services, Education, Environment, Poverty, Regional statistics, Living condition survey, Time use, Health statistics, Business statistics, Food balance, Tourism, Institution building			

* see Annex for details

As shown in Table 1, ISTAT has provided assistance to 14 countries in 24 fields of activity with large concentration as regards Social and Agricultural statistics, including Censuses. The estimation of the non-observed economy (NOE) is the assistance programme implemented by Istat with the largest number of countries. In addition to national programmes, Istat also led the Eurostat task forces for PECO and NIS for the exhaustivness of the National Accounts.
Among the different programmes carried out in these years, surely the Estimation of the informal sector in Russia, the Social Statistics projects in Romania as well as the Agricultural census and the on going Population Census in Albania represent the most relevant co-operation commitment Istat has been involved with.
What we would like also to highlight is that co-operation projects are of mutual benefit. In fact, although the involvement of competent staff willing to participate in co-operation activities has often been difficult and cumbersome, the experts themselves have generally been extremely satisfied with the work developed with colleagues of other countries and often recognise the usefulness from a professional point of view of the technical and comparative work to be developed in these occasions.

Table 2: *Number of man/days of ISTAT assistance by country, 1993 – 2001 (June)*

Countries	Man/days
Albania	777
Romania	470
Russian Federation	101
FYROM	70
Bosnia i Herzegovina	52
Ukraine	33
Latvia	23
Others (including co-ordination)	69
Total	1,595

4 Main results of specific cases

As just indicated in Table 1, over approximately eight years of activity Istat has worked with many different partner countries in several areas of intervention. Nevertheless, Table 2 clearly shows that the strongest links have been established with three countries: Albania, Romania, and the Russian Federation.
On the basis of the issues outlined in paragraph 2, such countries provide a good example of the approach followed within the three different group of transition countries considered: a candidate country (Romania), a new independent state (Russian Federation), and a Balkan country (Albania). Based on the policy context described, relationships with these countries have also resulted into three different co-operation approaches: demand to accession driven (Romania), purely demand driven (the Russian Federation) and strong bilateral involvement (Albania).

4.1.1 Romania

As highlighted by Table 2, Romania is the country with which Istat established since 1994 the most intensive relationships after Albania. Since the very beginning the two Institute of statistics agreed to work in the field of social statistics and poverty measurement, even if actions in other field have been carried out from time to time. In fact, as reported in the Annex, several interesting programmes have been implemented: a three years training course programme for trainees financed by UNICEF Romania, consisting of two main components, and a survey implementation financed by Phare for the harmonisation of Romanian surveys to European standards.
As regards the Unicef programme, the first component concerned the training of staff of the National Statistical Office and other ministries and research institutions, as well as staff of local authorities. The aim was to contribute to the dissemination of a statistical culture in the main public bodies concerned with social issues, at the national and local level. Such an objective was instrumental for a correct functioning of the statistical system and for the development of synergies between statistics producers and users, and between central and local institutions.
The second component of the Unicef Romania programme addressed the need to develop a training package for trainers for the dissemination of methodologies, information, scientific literature and the statistical culture to a large number of national and local staff concerned with social issues and statistics. Consequently, the training package developed by Istat, was conducted at two different levels: first a small group of participants was trained in Italy with the intent of creating a group of trainers able to transfer the know-how to larger groups in Romania, both at the national and local level. Then a second series of workshops were held in Romania by the staff trained in Italy assisted by Istat experts.
The final output of the training programme is a manual structured in modules thought to suit to the needs of users. The modules can be used all together and in the sequence proposed, or be selected and delivered according to specific needs. To this end each module contains the objectives and the instructions for trainers on how to adapt the material to the audience they address. After completion of the training programme, a volume "Social Trends", analysing all the main factors of the last 5 years Romanian social development has been written by the Romanian experts with Istat supervision.
The second important experience, started in 1999 and to be completed in these days (June 2001), is the support for updating the main statistical surveys in this area

according to the European standards: households budget, time use, living conditions, and health survey. The Statistical Office of Romania is now implementing these surveys using the same methodology and producing the same outputs as any other European country, which is particularly relevant for Romania for the discussions with the European Commission on the Accession agreement.

4.2 Russian Federation

A first Memorandum of Understanding was signed between Istat and the Goskomstat of Russia at the end of 1994, to study and exchange expertise in any statistical domain that would have been commonly agreed as being a priority one. A second agreement has been signed only two years ago, providing for co-operation in more specific areas of interest, such as business and regional statistics.

The first concrete project established since 1995, on the basis of mutual interest and in the framework of Tacis co-operation, has been the project to design a methodology for the measurement of the Non-Observed Economy (NOE). Efforts had already been made by the Goskomstat both in the analysis aimed at identifying sectors where the NOE is more likely to be present and in preparing solutions to the problem. The results were far from negligible. However, due to the unprecedented rapid growth of the NOE, Goskomstat's experts were looking for new ways to validate their NOE adjustments (and therefore the GDP estimates), also by referring to the international experience. The aim of the project was thus to contribute to ensure the exhaustiveness of the national accounts estimates of the Russian Federation, by cross checking the estimates already made by the Goskomstat with the results obtained with an alternative method based on the Italian approach. The project, which included a pilot test in a region, has been a success, and came to an end at the beginning of year 2000; upon request of Goskomstat, again in the framework of the Tacis programme, a new project on the estimation of the NOE will start in the fall of this year.

Another project established between Istat and the Russian Federation aims at investigating Individual Entrepreneurs. The project, funded in the framework of the Tacis programme and with a focus on strengthening regional statistics, will be mainly conducted in the Rostov region of the Federation.

An important meeting was also organised in the Russian Federation to present to all Tacis countries the Handbook on the Non-Observed Economy that OECD, in co-operation with other international organisation and a few National Statistical Institute – Istat and the CBS of the Netherlands – are preparing and that will soon be published: the Seminar provided a forum to present to all countries methods and best practices in the field of NOE estimation and was a good occasion to test a methodological text against users' requirements.

Other areas of intervention concern foreign trade statistics and other ad hoc consultancies on subjects of mutual interest.

4.3 Albania

The relationship with Albania started very soon after its opening in 1993, and grew very close over time.

A Memorandum of Understanding was signed, and initially activities started in different fields under the Phare Programme. As over time only two NSIs have been working with Albania, Istat and INSEE - and the latter has interrupted its activities approximately two years ago - it has been possible to follow the endeavours and the achievements of INSTAT, the Institute of Statistics of Albania, from a privileged viewpoint.

As mentioned, at the onset of co-operation, technical assistance activities were conducted in the framework of the Phare programme in several fields like labour force, cartography and agriculture. In the framework of a UNFPA funded project, Istat also hosted a demographer in a joint programme with the University of Rome.

In the field of agriculture, a lot of progress within Phare funded activities was already achieved in the planning of an Agricultural Census, whose funding however was still not available. In 1995, the Italian Government through the Italian Ministry of Foreign Affairs allocated the necessary funds to finish the planning and to implement the 1st Albanian General Census of Agriculture. The Census was finally implemented in autumn 1998, but details of this important programme will be given in this volume by our Albanian colleagues. However it can be stated here, that in spite of some procedural difficulties, the Italian project, based on a more continuous type of collaboration than the one usually carried out through Phare activities, proceeded smoothly and allowed the establishment of a fruitful and more continuous relationship between the experts of the two sides.

The involvement in the field of agriculture will continue with the implementation of the farm register and the design of a series of sample surveys to be carried out in conjunction with the Ministry of Agriculture and Food.

Following the successful implementation of the Agriculture Census, the Italian Government has committed additional funds for the implementation of the Population and Housing Census. This multi-donor programme has been planned and implemented by our Albanian colleagues with great engagement, especially taking into consideration that funds have become available very late, compared with the time normally required to plan such operations. Census took off as scheduled on last April 1st and data is now being transferred to the center and soon preliminary results will be available.

Italy and Istat are most hopeful and willing to continue their support to Albania: we hope to be able to further support census activities, so that after the successful completion of the Population Census it is possible to set the basis for the establishment of an Economic Census, and together with the statistical activities implemented with the support of other institutions (mainly international organisations and Sweden), INSTAT will have created and strengthened its basic statistical infrastructure which constitute the back bone of any modern statistical system.

Albanian Census of Agriculture Holdings

Myhjidin Llagami

Institute of Statistics, Rruga "Leke Dukagjini", No 5, Tirana, ALBANIA
e-mail; mllagami@instat.gov.al

1. Introduction

The General Census of households also called the General Census of Agriculture is part of the activities an Institute of Statistics should develop.
The big social-economic changes faced in our country, and especially in the Albanian agriculture after agricultural land privatization and creation of private households made the reflection of this new reality from the statistical point of view a necessity. This reality required the creation of the appropriate institutional structures and a new concept of the statistical activity in the agricultural field.
Under the increasing need for agricultural statistical data which are reliable and comparable to the European standards, with the initiative of INSTAT and Italian INSTAT and the consent of the Albanian Government, it was decided to conduct the General Census of Agriculture. It is a direct planned statistical action organized and applied on the basis of contemporary and scientific criteria and definitions. The economical environment where the Census conducted was totally different from the ones before, and the structures of the free market economy, though very fragile were the only orientation to determine the definitions. The Census was made possible through the technical assistance of Italian INSTAT, EUROSTAT and the financial support of the Italian Government.
The General Census of Agricultural Households aims at:

a) Assessing agricultural, forestry and livestock households in Albania.
b) Determine the principal structural characteristics of the agricultural, forestry and livestock households as well as the whole Albanian agriculture including the following aspects:
   - management system
   - juridical and organizing form
   - land area
   - way of tilling the land
   - farm practices
   - number of livestock
   - mechanical equipment used
   - households' labor
   - other aspects.
c) Developing the necessary data basis to conduct sampling surveys in the agricultural households in the years between two general Censuses.
d) Allowing the availability of information to a community and village level.

2. Legal basis

The General Census of agricultural households in Albania was based on several legal and sub-legal acts. In support to the Article 9 of Law No.7687, date 16/03/1993 “On Statistics”, the Council of Ministers drew up the Decision No.355, date 13/05/1996 “For the conduct the General Census of the agricultural households” concerning:

- Periodicity of accomplishment the general census of agriculture. The general census of Albanian agricultural households will be accomplished once in 10 years.
- The institutions involved for the general census of agricultural households is INSTAT, Ministry of Agriculture and Food, and Ministry of Local Authority.
- Establishment and functioning of Central Census Commission led by the Minister of Ministry of Agriculture and Food . The Central Census Commission determines the Census Commission in the districts and Census offices in the communities.
- INSTAT provides methodological leadership, data elaboration and editing of census results.
- The necessary expenditures to afford this activity are realized by INSTAT and the support of the Italian Government.
- Survey area of general census includes the entire agricultural live-stock and forest households (private or public).
- Private and public entities that are object of general census of agricultural households, are obliged to give correct and reliable information according to the questionnaire compiled by INSTAT.
- The data collected by the general census of agricultural households are statistically of confidential nature. Their distribution and edition can be done only in a grouped way.

In addition, in support to the Article 9 of Law No.7687, date 16/03/1996 “On Statistics” the Council of Ministers drew up the Decision No.34, date 17/01/1998, for a change on the Decision of the Council of Ministers No.355, date 13/05/1996 “For the conduct of the General Census of agricultural households”.
Implementing item 10 of the Council of Ministers No.355, date 13/06/1996, INSTAT and the Ministry of Agriculture and Food jointly compiled the instructions guide No.104 and 1765 which splits the tasks between the three institutions charged to conduct the Census: INSTAT, Ministry of Agriculture and Food and the State Secretariat for the Local Authority. In addition, this guide defined the composition of the regional commissions of the census depending from the Central Commission of Census approved by the above decision of the Government.
Implementing item 10 of the Decision of Council of Ministers No.355, date 13/05/1996, INSTAT and the Ministry of Agriculture and Food jointly compiled the guide No.1, date 21/01/1997 “On the progress made for the general census of agricultural census of agricultural household” which charged the registering commissions in districts and communes with concrete tasks.
Implementing the Decision of the Council of Ministers No.355, date 13/05/1996 changed by the Decision of Council of Ministers No.34, date 17/01/1998 instruction No.3, date 09/03/1998 was drawn up jointly between INSTAT, the Ministry of Agriculture and Food and the State Secretariat for the Local Authority. This guide precisely defined the tasks of the three institutions and the organs of registration in the

field. In addition, it defined the composition of the census commissions from the center to the field, and it approved the calendar of census activities.

3. Pilot survey

After designing the census questionnaire and the methodology of data collection, in March-April 1994 a pilot survey for testing the questionnaire was conducted in 435 private agricultural households in 6 villages: Shakuj and Bregas – Lushnja district; Mamaj and Cerril – Tepelena district; Lurth and Katunnd i Vjeter – Mirdita district. The Ministry of Agriculture and Food was a close cooperator to collect data and it involved 6 specialists (1 for each village) as enumerators. Testing aim was the verification of questionnaire and interview functioning but also to find out census organisational aspects. Testing was also used to determine the assimilation level of instruction by enumerators and to determine the number of household that the enumerators on the average can interview during a day; the questionnaire number that can control the community controller; daily questionnaire number that an operator can work on; functioning of computing program to make data elaboration and controlling. The test was organized to make a fair prediction according to census expenditures, enumerators number, controller number of districts and communities, operator's number for data elaboration, time limit of data collection phase (agricultural households questionnaire) of computing elaboration of collected material etc.
After completing the test, some simple tables were compiled containing data from the test. Some improvements were made to parts and questions of the questionnaire where problems were spotted during the test.
The purpose of this testing was to verify the functioning of the questionnaire and the interview as well as testing the organizing parts of the census.

4. Households lists

Until '90, agricultural production in Albania was based on the activity of 1000 cooperatives and agricultural enterprises. In the agricultural sector like in all the other sectors of our economy, private property didn't exist. According to law "For Soil" No. 7501, dt.19. 07.91 approved by the Albanian parliament, the agricultural soil was distributed to those families forming part of the former agricultural cooperative, according to the number and employees of former agricultural enterprises. As a result, private agricultural households were created.
To accomplish the general census of agricultural households, it was necessary to use the compilation created after ’90. It will serve as a base for enumerators while testing agricultural households.
Before beginning with the compilation of these lists, we first accomplished codifications for districts, communes municipals and the entire villages, something that had never been done before. During March-July 1996, INSTAT in cooperation with MOAF organized and compiled coded lists of agricultural households in country. To compile these lists, the documentation of agricultural land distribution found in Land Registry offices was used.

Households lists contain districts, community/municipal code and the village code where the households is included. The agricultural households lists contain data for its households manager (name, surname, father name) and the surface of agricultural soil shown in m2.
Also in this list are represented the changes that occurred from the moment of soil distribution (1991) up to lists compilation (1996).
The compilation of the agricultural households lists was accomplished by the employees of civil status in the communities. Based on the compiled lists it appeared that there were around 455.000 agricultural households to be interviewed. Because of the time between list compilation and the interview two years later, the migration of the population during that period, and splits or joinings of the households it was necessary to update the list of the households. According to the criteria defined by INSTAT, this process was realized during May and mainly in June 1998 during the interview period.

5. Organizing structure

The organization of the census is based on the input of the following organisations:

5.1 The Census Central Commission

The Census Central Commission was created by the Decision of the Council of Ministers No.355, date 13/05/1996. This commission determines the composition of the local commissions and the census offices in communes or municipalities. The Census Central Commission is directed by the Minister of Agriculture and Food.
The Census Central Commission is the highest organism managing and organizing the conduct of agricultural census. This commission approves the census methodology including the questionnaire, instructions for collecting the data and the final results of the census.

5.2 The Census Commission in the district

The census commission in the district is directed by the Chairman of the District.
This commission manages and organizes the work at a district level. It is responsible to:
Assist the communes under its jurisdiction to correctly and accurately perform the functions charged by the superior census organs; popularize the purpose of the census; pay attention that the questionnaires get completed on time and accurately in communities according to the calendar compiled for this purpose; provide technical assistance to the census offices in communities; train the community supervisors throughout the district supervisors; provide working environment (office) for the members of the commission.

5.3 Census Office in the Community/Municipality

Census Office in the Community is directed by the community chairman.
Census office in the community is responsible to:
Precisely implement the tasks imposed by the superior census organs; select and recruit the enumerators based on the defined criteria; organize the conduct of interview in the villages to be covered and pay attention that the questionnaires are completed accurately and in time; implement the training of the enumerators in the villages by means of the supervisors; after training, incapable enumerators are replaced; provide working

environment (office) for the members of the census commission in the commune; receive the census materials from the district (directorate of agriculture, directorate of statistics); collect the completed questionnaires and forward them to the district center; complete the tables with the preliminary results of the census (table FR-1 and FR-2) and forward them to the census commission in the district.

6. Household unit (survey)

The general agricultural census household unit is the agricultural, forestry and livestock household. In the concept **agricultural, forestry and livestock household for census effect we include the techno-economical aspect under a unique management system made of the land, though in non-continuous parcels where the agricultural, forestry and livestock production is realized, it is done by a manager, that is a physical person (a single person or group of persons), "state institution" or "others".**

The essential distinctive characteristics of the agricultural households are:

a) Land use for agricultural, forestry and/or livestock production. Land may consist of one or more parcels.

b) Management is performed by one manager (owner). Organizing and juridical form can be: "a physical person", (single person, or group of persons).

7. Census population

The Census population includes all the units that are included in the definition of the agricultural, forestry or livestock household regardless of size.

Excluded from the Census Population are:

a) Land of agricultural households not used for agricultural, forestry or livestock production, in other words, when land is used for other purposes.
b) Agricultural households totally abandoned (immigration, left the district and other causes) though land might spontaneously produce products.
c) Agricultural land distributed by villages or communes commissions, but refused for different reasons by families.
d) Agricultural land not distributed yet.
e) Livestock number or fruit tree pastures located around private buildings in the cities.
f) Fruit trees, and especially olive parcels not distributed yet, or distributed to the former workers of Agricultural Enterprises, or to the owners as a result of the remote location of their dwelling unit.

8. Personnel training

Special importance was attached to training personnel involved in agricultural households census. The reason of it was the census, requiring a high number of

enumerators. Also, it was the first time that such an undertaking took place in our country.
This phase included the staff (6 in numbers) of the agricultural sector of INSTAT training district controllers (120 persons) community controllers (910 persons) and villages enumerators training. (3450 persons).
In cooperation with the Italian experts a detailed calendar for the training of controllers and enumerators was prepared. This plan was based on the methodological book "Instruction about data collection" prepared in advance. This contained an explanation about theorical aspects of census operations and practical examples about questionnaire completion, the connections between different categories, correction of mistakes etc.
4-9 May 1998 a group training of district controllers was divided in 2 groups. For 3 days each group was trained in Tirana. This training was established with theorical explanations of census procedures, practical examples of filling in questionnaires, correction of mistakes, and a practice session with an agricultural households.

9. Data collection

According to the approved calendar, the household interviewing process started on June 1, 1998. This census was accomplished through a direct interrogation to the manager (owner) of an agricultural household. According to the agricultural household definition and survey area, the general agricultural census was carried out in 466.809 agricultural households (private and public), 2968 villages and cities, 368 communities and municipalities, 36 districts and 12 prefectures.
The enumerator, in accordance with the previous households list sent by INSTAT, completed the questionnaire according to the data the households manager gave him (in case he was absent, another family member capable of giving this information was doing it) by means of a direct interview.
The data that were object of the census were collected through the questionnaires and forms prepared by INSTAT. It was not allowed to use different models.
According to the categories and questions that the questionnaire contains, the data collected refer to the agricultural year 1997-1998 or 1 June 1998.
Specifically, this concerns the data for management form, territory use, work force, utilization of mechanical devices, production selling, soil watering and fertilization, refer to the agricultural year 1997-1998.
The data for organizational and juridical form, total surface and agricultural surface actually used, the number of livestock, refer to the agricultural year, 1 June 1998.
For data collection 4500 persons have contributed (67% of which with a high level of education). These persons had the function of: enumerator, community and district controller and member of census commission, etc.
On average, each enumerator took care of about 135 households and the average daily work consisted of 5-6 interviews. That's why the interviewing period was determined to last for 30 days, 1-30 June. This time limit was respected and the interviews were completed on the first week of July on a country level.
During that phase, assisting and quality control missions were organised in some districts of the country with the help of the specialists of INSTAT, giving solutions to problems that emerged. In addition, an INSTAT employee was in contact through the telephone with the districts during the entired period.

This phase was combined with an advertising campaign broadcasting a television spot, by organizing some television programs in the studio and in the field in the districts of Kruja and Durres discussing about the Census and its purpose. A special poster prepared for the purpose of the agricultural household census was distributed and displayed in all communities. The census was also advertised through the local electronic and written media.
The period July – August served to check the questionnaires by the communities and district supervisors, to complete the “FR” tables with the preliminary census data, to collect the materials from the communes to the district centers and then forward them to INSTAT.

10. Data entry processing

Initially, this process began with adopting a data entry hall in the environment of INSTAT, by installing an information network based on a contemporary hardware and software technology. Computers used were IBM PC Pentium Windows 95, Novel 4.11 Server and a NT 4.0 Servers. The information network had a communication speed of 100 Mhz.
The program used for data entry was designed by an Italian specialist, programmer in Visual Basic 5.0. The information file format is a Database and the statistical data processing is in SPSS format. The application is of a Client – Server level, considering the accurate data entry and the continuity of work in every aspect of network communication.
The Data entry program incorporated more than 100 checking procedures defined by INSTAT and Italian statisticians. Through these checks it was possible to provide a checked data registration and data saving in the original and checked format spotting out the differences for study purposes. It has to be emphasized that the procedure used in this census is in the experimenting and application stage in the analogous European institutions.
After installing 40 computers in the information hall, 80 computers operators were recruited, who went through an interviewing process, and worked in two shifts for around 5 months. Monitoring of operators work was performed by four supervisors, and the whole process was followed step by step by the INSTAT working group.

10.1 Leading Type (Leader - Labor Ratio)

From the data, it appears that there are 466.809 agricultural economic units, out of which 90.5 % belong to “ the family labor group only”, this being explained by the fact that agricultural economic units have a limited space, and the leader together with the unit members cope with almost the whole volume of work; 8.3 % are included in “family labor dominant” group, and 1.2 % are included in “labor outside the family dominant” group.
From the above stated number of agricultural economic units registered in our country, only 93 economic units belong to “state entity” group, whereas the rest belong to “ physical person” group, being private agricultural units. For this reason during the analyses of the features, the bulk part refers mainly to “physical person economic units” group.

10.2 Land Ownership

According to the data at our disposal, it appeared that 92.3 % of agricultural economic units own the land, and this property represents 95.9 % of Usable Agricultural Land (UAL); 3.4% use the land which represents 1.3% of UAL, and only 4.3 % of economic units use the land with rent and other combinations, owned and in use, owned and rented; in use and rented, owned, in use and rented.

10.3 Way of land tilling

When speaking about the way of land tilling, we mean the way of basic land tilling. 25.8 % of economic agricultural units till the land manually, 53.6 % use animals, and 53.6% use mechanical means. From the data it results that the number of economic agricultural units using mechanized means increases with the increase of UAL types.
In some north-eastern districts, because of the mountainous terrain which make it difficult to use mechanized means, the number of economic agricultural units which till the land in a primitive way (manually or with animals) ranges between 75%- 89%, whereas in flat areas the above figure goes down to 12 %.

10.4 Land size

The total land size of registered agricultural economic units is about 1.889.497 ha, of which UAL represents 43% ; forestry represents 54% and unusable agriculture land and non-agricultural land represents 3%.

10.5 Economic units cultivating crops

The number of economic units cultivating crops amounts to 439.732, representing 94.2% of all the country economic units. Economic units cultivating bread grains represent 84% of the number of the units cultivating plants. This is explained by the fact that the bulk part of economic units living in the countryside secure bread by themselves. Economic units cultivating patatoes and beans respectively represent 36 and 55 % of the number of economic units cultivating plants, whereas 72% of the economic units cultivate vegetables. Economic units cultivating industrial plants represent only 9 % of the number of economic units. For plants such as sunflower, sugar beet and soybean the number of economic units represent 2%, and cultures like cotton and rice which were formerly cultivated, at present are no longer planted. This is explained by the fact that in our country many of the factories existing before 1990 and processing agricultural products are no longer in use.

10.6 Land culivated with plants

The data for the cultivated land for each culture refer to July 1, 1997- June 30, 1998 agricultural year. Taking into account the limited land size of agricultural economic units, in order to be as accurate as possible, we came at the conclusion that the data referring to land be given in square meters. There are only data for the land cultivated with “main crops” and “secondary crops”.
According to the data at our disposal, the land cultivated with crops represents 42.6% of usable agricultural land, equal to 345.258 ha. Bread grains have been cultivated on 173.406 ha. or 50.2%. Within the group of bread grains, the bigger specific part is taken by wheat with 66.7 %, and maize with 26%. An agricultural economic unit cultivating bread grains has averagely planted 0.47 ha. with bread grains.

Following the bread grains which occupy 50.2 % of cultivated land, the bigger specific part is occupied by fodder with 27%. Other crops like potatoes, vegetables, and beans take 14.2 % of the land, whereas 8.6% is land laid wast.
Considering the number of the crops cultivated in our country like wheat, maize, beans, vegetables, potatoes, etc. we come to the conclusion that although the average size of private agricultural economic unit is limited, farmers work hard to secure a good part of production by their agricultural economic units.

10.7 Economic units using irrigation and fertilizers and their respective size

By analysing the data at our disposal, with regard to economic units using irrigation, the conclusion is reached that their number and the irrigated space they comprise is rather limited. The data for the main cultures are as following:

In percentage

	number	size of land
Bread Grains	37.4	14.5
Potato	41.5	39.1
Bean	44.2	27.5
Vegetables	68.9	54.6
Fodder	36.4	26.1

On national scale it appears that only 22% of the cultivated land is irrigated.
There is almost the same situation as far as the use of chemical and organic fertilizers is concerned. The land fertilized with chemicals and organic fertilizers is respectively about 63%, representing 15 % of the cultivated land.

10.8 Economic units and the number of fruit-trees

The data for fruit-trees, different from the data for cultivated crops have been reported in the number of fruit- trees and in square meters.
There are 328.283 economic units owning fruit-trees, equal to 70.3% of agricultural economic units on national scale. 87 % of economic units which have fruit cultures have fruit-trees. For olives and citrus the figures are respectively 34% and 9 %. The space occupied by fruit-trees cultures reaches 33.860 ha, representing only 4 % of UAL. There are about 3.1 million fruit- trees, 3 million olives and 243 thousand citrus trees.
Starting from the average number of economic units owning fruit-trees, the conclusion is reached that fruit-trees are mostly spread and do not exist in blocks. This fact has been taken into account also when the questionnaire was being drafted, when alongside the number of fruit-trees (in case they were spread) the leader of the economic unit would be asked for data for the space (in square meters) that the fruit-trees covered. In case the leader was not able to provide data for the space, then the interviewer, basing himself on “the book for collecting data”, would convert the number of fruit-trees into land space according to the following criteria:

For fruit-trees	1 tree = 35 square meters
olives	1 tree = 80 square meters
citrus	1 tree = 30 square meters

The vineyard space is reported in square meters without making any conversion, whereas for vines the number of vine roots is reported.

10.9 Meadows and pastures
The space occupied by meadows and pastures is 430874 thousand ha, equal to 53.2 % of UAL. 3% belongs to private agricultural economic units, whereas 97 % pertains to economic units of the "State Entitities" group represented by meadow and pasture enterprises.

10.10 Forestries
In our country forestries occupy a considerable space of 1 million and 26 thousand ha, equal to 54 % of the total space. Only 0.04 % of this space belongs to private economic units, thus almost the whole space pertains to the "State Entities" economic type represented by forestry enterprises.
The agricultural economic units of the "State Entities" type are big enough and their average space reaches up to 18 thousand ha, this because the privatization process of meadows and pastures and also of orestries had not yet started at the moment of registration.

10.11 Animal Husbandry
The data for the situation of animal husbandry refer to June 1, 1998, agriculture registration starting day, when information was collected for the number of cattle according to gender, age and economic destination.
From the data collected, it turns out that 68% of the registered economic units cross-breed cattle; 24.2 % cross-breed sheep, and 14 % cross-breed goats.
Always referring to the cattle and sheep the number reaches to 635 thousand cattle; 1.5 million sheep and 808 thousand lamb. On average, an agricultural economic dealing with cross-breeding owns: two heads of cattle, 14 sheep and 12 goats.
Referring to the herds structure data, it results that the number of cows in production represent 58% of the total number; the sheep 74% and the goats 70 % of the total.
In analysing the geographical distribution of the main breeds of animal husbandry, the conclusion is reached that cattle is almost equally distributed through out the country. The situation is quite the opposite for goats which are concentrated in some places. Thus, only 3 prefectures out of 12, own 50 % of the sheep and goats. They are located in the northern, central and southern part of Albania. Whereas pigs, though limited in number, are to be found in the northern part of the country, and two prefetures own 90% of the total number.

10.12 Sale of production
As above stated, a good part of agricultural economic units produce for their consumption and the specific contribution of their articles is rather small. Thus, only 51.5% of the economic units trade agricultural and dairy products. In the flat area of the country where the average space of economic units is bigger than 1.5 ha, the above figures range between 69-73 %.
In the northern part where the size of economic units is rather limited (less than 0.5 ha), this figure goes down to 25%. 37% of economic units trade animal husbandry products. 35 % trade agricultural products, and only 15% trade processed products.

10.13 Labor

The data for labor refer to persons who at the moment of registration are 15 years of age or more.

Analysing the economic units according to the number of members and labor days working in economic units, the data are as following data:

	In percentage	
Number of members	number	labor days
with 1 member	6.6	2.3
with 2 members	46.0	31.9
with 3 members	18.9	19.7
with 4 members	17.2	23.9
with 5 members	6.9	12.0
with more than 5 members	4.4	10.0

Data referring to the leader of economic unit, spouse and other members are as following:

	In percentage		Absolute
	Number	labor days	labor days
Unit leader	34.8	42.2	149
Spouse	30.3	29.6	120
Other members	34.0	28.2	99
Total number	100.0	100.0	123

Registration data referring to age-groups are as following:

	In percentage	
Age-groups	Number	labor days
up to 19 years	8.8	7.1
20 - 29 years	22.7	19.6
30 - 39 years	23.5	26.2
40 - 49 years	17.4	21.0
50 - 59 years	12.6	14.4
60 - 64 years	5.5	5.0
more than 64 years	9.3	6.2

Regarding the educational level of the family members of the leader of the economic unit, the data are as following: 5,7 % are illiterate; 13,7 have finished 4 years of elementary education; 56.7% have finished 8-grade education, 21.9 % have concluded secondary eucation, and only 2 per cent are graduated.

8 June 2001

SATELLITE MEETING

Organised by FAO

FAOSTAT – The FAO Corporate Database

Abstract: FAO data, presented through FAOSTAT, are available in the FAO Web site: www.fao.org . Databases are organised in form of a table where columns indicate notes, domains, data collections and last updates and rows provide further details on data sets included in each data domain. For example, the Agricultural Production domain provides data-sets on crop primary, crops processed, etc. and Means of Production gives details about agricultural machinery, fertilisers, etc. Individual data series can be accessed by double clicking on the icon referring to the subject chosen and methodological information can be read by double clicking on the yellow icon on the left-hand side of screen. A sample view of the page can be seen in the Annex 1 given on the last page. It may be mentioned here that the FAO classification covers more than 700 food and agriculture commodities relating to crops, live animals and animal products. In addition to primary products, a large variety of derived products are also covered. These items are important for determining food availability. FAO definitions of individual commodities can be found in the document 'Definition and Classification of Commodities' and in the FAO Web site. Annex 2 gives an example of questions most frequently asked by database users. Some of the important domains are described below.

1. Agricultural Production

This domain provides time series data on production for all agricultural products. For primary crops harvested area, yield, production and seed are presented. For selected derived crop products, production is shown. For live animals, stock numbers are shown. For primary livestock products such as meat and milk, production is shown, as well as the input to that production, e.g. the number of animals slaughtered, carcass weight, etc. The time reference for statistics on area and production of crops is the calendar year. That is to say, the data for any particular crop are reported under the calendar year in which the entire harvest or the bulk of it took place. This does not necessarily mean that for a given commodity the production data are aggregated month by month, from January to December, although this is true for certain crops such as tea, sisal, palm kernels, palm oil, rubber, coconuts and, in certain countries, sugar cane and bananas, which are harvested almost uniformly throughout the year. The harvest of other crops, however, is generally limited to a few months and even, in certain cases, to a few weeks. Production of these crops is reported by the various countries in different ways: by calendar year, agricultural year, marketing year, etc. Whatever the statistical period used by the countries for presentation of area and production data, these data are allocated, commodity by commodity, to the calendar year in which the entire harvest or the bulk of it took place. Obviously, a crop that is harvested at the end of the calendar year will be utilised mostly during the year following the calendar year under which the production figures are reported. It should be noted that the adoption of a calendar-year time reference period inevitably means that, in a number of cases, crops assigned by country to a particular split year may appear under two different calendar years.

Livestock numbers have been grouped in 12-month periods ending 30 September or 30 December in the tables. For example, animals enumerated in a given country any time between 1 October and 30 September of the following year are shown under the latter year. As regards livestock products, data on meat, milk and milk products relate to calendar years, with a few exceptions. Data for other animal products that are produced only in certain periods of the year, for example, honey and wool, are allocated to the calendar, following a policy similar to that adopted for crops.

1.1 Crop areas
Figures for crop areas generally refer to harvested areas, although for permanent crops data may refer to total planted area.

1.2 Yields per hectare
All yields per hectare are given in hectograms. In all cases, they are computed from detailed area and production data expressed in hectares and metric tons. Data on yields of permanent crops are not as reliable as those for temporary crops either because most of the area information may correspond to planted area, as for grapes, or because of the scarcity and unreliability of the area reported by the countries, as for example for cocoa and coffee.
The Agricultural Production domain is presented in five data collections:
i) Crops Primary - This group contains all major crops, which include the cereals, pulses, roots and tubers, sugar crops, oil-bearing crops, fruits and vegetables. In addition this group contains nuts, fibre crops, spices, stimulant crops (coffee, cocoa, tea), fodder crops and other crops (tobacco, rubber).
ii) Crops Derived - This group contains derived products of vegetal origin. Their parent products, are found in the group Crops Primary.
iii) Live Animals - This group contains the classification of animals, large (e.g. cattle) and small (e.g. rabbits).
iv) Livestock Primary - This group contains primary animal products. These are the products of slaughtered animals, such as meat, offal, slaughter fats and hides and skins, and the products of live animals, such as milk, eggs, wool and honey.
v) Livestock Derived - This group contains derived animal products. These are the various forms of processed meat, rendered fats, skimmed milk, butter and cheese. Their parent products are found in the group Livestock Primary.

2. Agriculture and Food Trade domain

This domain contains the most comprehensive and detailed time series data on foreign trade in agricultural and food commodities. It provides the value of foreign trade, expressed in US$, as well as the volumes of trade. The latter feature makes this set crucial in determining the quantities of food supply in individual countries. In general, trade data have been supplied by governments through magnetic tapes, national publications and FAO questionnaires. In particular, for EC member countries, with the exception of Spain, data obtained from EUROSTAT have been used. In addition, maximum use has been made of the magnetic tapes provided by the United Nations Statistical Division. To make the coverage as complete as possible, official trade data have sometimes been supplemented with data from unofficial sources. Use has also been made of trade information supplied by other national or international agencies or organisations. In the absence of reliable sources or when information for the latest year is not available in time, figures for quantities and values are estimated on the basis of trade returns of trading partners. In a few instances, when information is available in terms of quantities only, corresponding values are estimated, using unit values based on data from trading partners.

2.1 Classification used

In 1988, many countries adopted the Standard international trade classification, Revision 3 – SITC, Rev. 3 (Statistical Papers, Series M. No 34, Revision 3, Statistical Office of the United Nations) or the Harmonised commodity description and coding system (HS) of the Customs Co-operation Council, which is one to one correlated to the SITC, Rev. 3.

3. Commodities Balances

This data domain presents the most comprehensive and detailed data on agriculture. It provides the flow from production and trade to the utilisation of primary and derived food and agricultural commodities. The commodity data are available in the form of Supply Utilisation Accounts (SUAs) which present consistent and interlinked information. Total supply of a commodity (equal to production and imports and the off take from stocks) matches total utilisation (i.e. exports, additions to stocks, feed, seed and food use, quantities processed and other utilisation) of the commodity. Chains of processing (e.g. wheat to flour of wheat to bread) are accounted for through a specific utilisation category (quantities processed).

Production and utilisation of some derived products, generally at the second and higher stages of processing, may not always refer to a country's actual total. This occurs whenever there is net exports of these derived products. In such cases, to account for the use of the originating commodity (say wheat flour), and to decrease its availability in the country, a sufficient quantity is allocated to produce the derived product (say bread), which is subsequently exported. Consequently, the quantity shown as the country's production of the derived product is just to cover the net export. Net imports of such derived product will be allocated to the appropriate utilisation category.

4. Food Balance Sheet

FAO's Food Balance Sheets present a comprehensive picture of the pattern of a country's food supply and utilisation during a specified reference period. They show for each food item, i.e. each primary product and a number of processed products, the sources of supply and its utilisation. The total quantity of foodstuffs produced in country, added to the total quantity imported, and adjusted for any change in stocks that may have occurred since the beginning of the reference period, gives the supply during that period. On the utilisation side a distinction is made between the quantities exported, fed to livestock, used for seed, put to manufacture for food use and other uses, losses during storage and transportation, and food supplies available for human consumption. The per caput supplies of each such food item available for human consumption is then

obtained by dividing the respective quantity by the related data on the population actually partaking of it. Data on per caput food supplies are expressed in terms of quantity and, by applying appropriate food composition factors for all primary and processed products, also in terms of caloric value and protein and fat content.
Annual food balance sheets tabulated regularly over a period of years will show the trends in the overall national food supply, disclose changes that may have taken place in the types of food consumed, i.e. the pattern of the diet, and reveal the extent to which the food supply of the country, as a whole, is adequate in relation to nutritional requirements.

5. Mean of Production Domain

This includes the following databases: fertilizers, agricultural machinery, pesticides.

5.1 Fertilizers

This chapter of the means of production domain provides time series data on production, trade, consumption of nitrogen, phosphate and potash fertilizers in terms of nutrients and prices in local currency for the same group of fertilizers. Time series data go from 1961 to 1999 and refer to all countries of the World and, whenever is possible, to country aggregates. Data generally refer to the Fertilizer Year 1 July-30 June and, in case they are referring to the calendar year, they are shown under the fertilizer year that begins in that calendar year (e.g. 1999 data are under 1999/00).

5.2 Agricultural Machinery

This database presents data on number in use as well as trade values of Agricultural Tractors, Harvester and Threshers and Milking Machines. Time series is available starting from 1961. For some countries information on number in use is poor and estimates are based on various materials such as data on shipment or data on imported numbers.

5.3 Pesticides

There are two separate databases:

1 "Pesticide Trade" follows the structure of the other databases. Data series is available from 1961 and refer to value of trade, covering insecticides, fungicides, herbicides, disinfectants and others. Definitions of the products are in compliance with definitions given in Standard trade classification systems (SITC, HS).

2 "Pesticide consumption" which is a particular one. Data collection is presented starting in 1990; time series and country coverage are incomplete due to a high rate of non-response. Therefore, in this domain continental aggregates are not feasible. Lists of countries for which data are available as well as explanations on the data are available in the notes section. The database refers to the quantity of pesticides used in or sold to the agricultural sector expressed in metric tons of active ingredients. However, it may happen that countries report data in formulation weight (including diluents and adjuvants) without specific indication.

6. Land Use

This domain presents data on land use and irrigation.

6.1 Land use

This database presents statistical time series starting in the year 1961. The land use categories are classified in connection with the FAO's recommendations for censuses of agriculture. Definitions of the land use categories are available in the notes section of the domain. The FAO land use statistics are based on point estimates derived from data collected in periodic agricultural censuses and missing years are estimated using a variety of information and data sources. Problems that are faced in collection, compilation and presentation of data on land use are mainly concerned with the concepts and definitions. In fact, certain item definitions of land categories which are adopted by various countries are different from those given by the FAO. For example, in the case of Permanent pastures and Woodland the dividing line between these categories is rather indefinite, especially in the case of shrubs, savannah, etc., which could be reported under either of these two categories. This is why the Forest and Woodland category is not presented in the FAOSTAT from 1995 onward and data for this element can be obtained from the FAO Forest Resources Division.

6.2 Irrigation

Data on irrigation are also available from 1961 and relate to areas equipped to provide water to the crops. These include areas equipped for full and partial control irrigation, spate irrigation areas, and equipped wetland or inland valley bottoms. This database in periodically reviewed in collaboration with FAO Land and Water Division which provides country reports and specific studies on the subject.

7. Producer Prices

This domain presents one series of data on prices received by farmers for the sale of their products.

7.1 Prices in local currency

This database present statistical time series starting in the year 1961. The last update dates back to 1995. Prices refer to primary products and the definitions used match those of the other domains. "Producer Prices" are determined at the farm gate or first-point-of-sale transactions when farmers participate in their capacity as sellers of their own products. A word of caution may be added for the users of this data set. Concept of the data on producer prices received in FAO might not always refer to the same selling points depending on the prevailing institutional set-up in the countries. Also different practices prevail in regard to individual commodities. Methods of arriving at national averages differ from one country to another. A comparison of data among countries therefore should be considered with these limitations in mind. Most of the data originated from country sources received either through the Agricultural Price Statistics Questionnaire or from official country publications. In 2002, FAO Statistics Division will start again to collect price data and will update this domain.

8. Problems of compiling internationally comparable agricultural statistics - The Large Volume of Data

A major difficulty faced by ESS arises from the enormous data that are required to be manipulated, and the multiplicity of the sources reporting the data. There is a huge quantity of annual figures relating to various elements of the supply utilisation accounts (SUAs) to be collected or estimated and then evaluated and recorded. Two examples: just the element in the accounts (031)- area or input - includes about 25000 country entries for a given year; some 5500 of them relate to harvested area of crops, while most of the other 20000 relate to "input", "slaughtering" and "animals producing".

The magnitude of the figures - in terms of units - moves between a few thousands hectares (HA) harvested of various crops in many countries and 42 million HA of rice harvested in India.

Element 051, Production/Output, includes more than 28500 country entries per year, of which about 8000 refer to production of primary crops and production of those few crop processed products shown in the FAO Production Yearbook. Production figures moves between a few thousand tons of various crops produced in many countries and more than millions of tons produced in some countries; for example, 280 million of tons of sugar cane in Brazil and 250 million in India; 255 million of tons of maize in the United States; 190 million of tons of paddy in China; 150 million of tons of maize for silage in the Russian Federation.

9. Data sources for FAOSTAT

a) Primarily, FAO Questionnaires. These are tailored questionnaires and sent to member countries of FAO;
b) national publications available in the ESS Library. There are some 3700 titles in all, including general yearbooks, agricultural yearbooks, pocket yearbooks; periodicals (quarterly, monthly, weekly, daily, etc.); early warning system correspondents' reports; FAO Country Representatives' reports;
c) country visits by ESS statisticians involving discussions with national experts;
d) international publications.

In some cases, official figures released by the countries can be found in these publications long before they reach our office through questionnaires or national publications. It is not always easy to ascertain which one of the various figures reported by the various sources is the most recent or the most reliable, although the national Yearbooks data are supposed to be "the best ones". However, they report historical data for a number of years with the last year shown being one to three years behind the current year.

It is always necessary to evaluate the data for consistency within the country and between countries, and to make sure that definitions and coverage of the data conform with FAO recommendations. To do that, it is indispensable that those in charge of collecting data have a fairly good knowledge of the various commodities they handle every day, from the point of view of agro-economic and biological characteristics. For example, they must know the difference between flax retted and flax scutched and hackled, between sown and harvested area, between dextrose and isoglucose, fruit juice and fruit nectar, carcass weight, dressed carcass weight and ready-to-cook weight, etc.

They should have a good idea of the possible and actual yield per hectare of various crops in different countries, take-off rate of livestock, potential and actual yield of milking cows, laying hens, etc.; the average weight of edible offal and slaughter fats of various species of animals, etc. Short descriptions of the commodities/elements can be found in the "Introduction", "Explanatory Notes" and "Notes on the tables" of the FAO Production Yearbook. Also, in the recently issued booklet "Definition and classification of commodities". Other related pages on this site are also very useful in this regard.

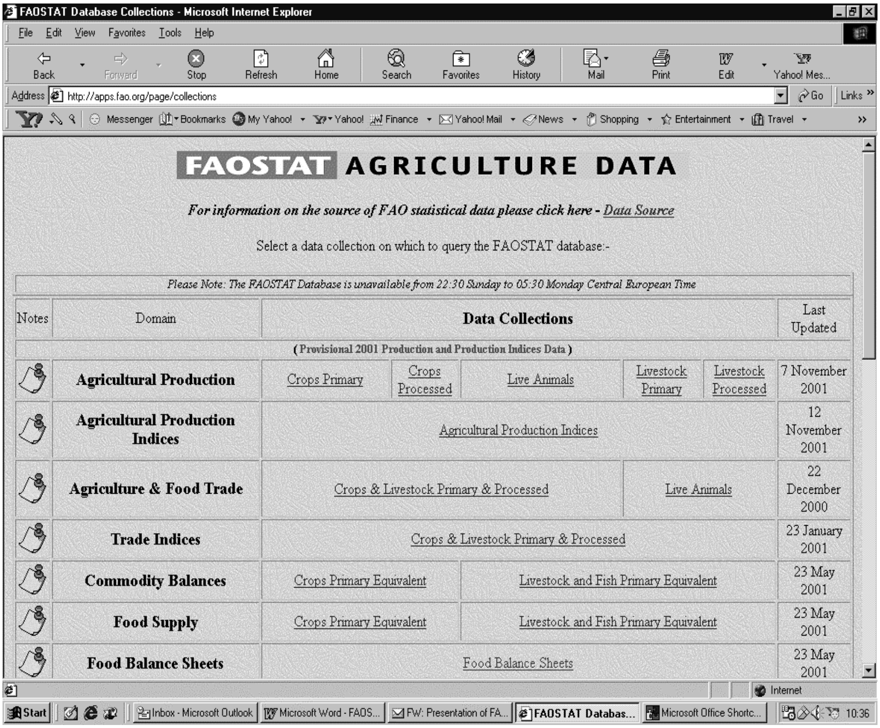
10. Estimation of Missing Data.

Although there are many sources of data, it should be noted that most sources deal with cash crops or with crops and livestock production which can influence the international markets. Production of crops and livestock, whose main or exclusive use is the subsistence of producers, is seldom covered by these sources; in other words, there is too much information for certain commodities and too little for others. As a result, there are data missing as well as gaps in the series.

In the absence of any figure from any official or unofficial source, estimates for missing data of area and production of a commodity can be made on the basis of various assumptions. The safest one appears to be following the observed trend of the commodity in question in the previous years, with an eye on the behaviour of that commodity in neighbouring countries.

The estimation of a complete time series, when neither official nor unofficial data are available, presents more difficulties. When we know from various sources, e.g., Trade Reports, Food Balance Sheets, Households Surveys, Census reports, etc., that a certain crop or a certain livestock product is produced in a given country, some (rough) estimates can be worked out on the basis of the little information available, taking into consideration the situation of this commodity in neighbouring countries with similar agro-economic conditions and similar food habits. Contacts should be established with the countries concerned, directly or through FAO or UN Officers living or visiting the countries, so that at least some first-hand information is made available to ESS. All this, of course, when the matter deserves such attention.

ANNEX 1


FAOSTAT Database Collections - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Yahoo! Mes...
Address http://apps.fao.org/page/collections
Go Links
Messenger Bookmarks My Yahoo! Yahoo! Finance Yahoo! Mail News Shopping Entertainment Travel
FAOSTAT AGRICULTURE DATA
For information on the source of FAO statistical data please click here - Data Source
Select a data collection on which to query the FAOSTAT database:-
Please Note: The FAOSTAT Database is unavailable from 22:30 Sunday to 05:30 Monday Central European Time
Notes
Domain
Data Collections
Last Updated
(Provisional 2001 Production and Production Indices Data)
Agricultural Production
Crops Primary
Crops Processed
Live Animals
Livestock Primary
Livestock Processed
7 November 2001
Agricultural Production Indices
Agricultural Production Indices
12 November 2001
Agriculture & Food Trade
Crops & Livestock Primary & Processed
Live Animals
22 December 2000
Trade Indices
Crops & Livestock Primary & Processed
23 January 2001
Commodity Balances
Crops Primary Equivalent
Livestock and Fish Primary Equivalent
23 May 2001
Food Supply
Crops Primary Equivalent
Livestock and Fish Primary Equivalent
23 May 2001
Food Balance Sheets
Food Balance Sheets
23 May 2001
Internet
Start
Inbox - Microsoft Outlook
Microsoft Word - FAOS...
FW: Presentation of FA...
FAOSTAT Databas...
Microsoft Office Shortc...
10:36

ANNEX 2
Frequently asked questions by the FAOSTAT Users.

Q 1: Where can I find data on Agriculture, Fishery, Forestry, etc?
A: Most of FAO Statistical data are available in the FAOSTAT located at the following site: http://www.fao.org/
by clicking on http://www.fao.org/ SELECT THE PANEL (Statistical Databases) AND CLICK, All databases will appear or directly go to the following site:
http://apps.fao.org/cgi-bin/nph-db.pl.
Most Domains within the FAOSTAT system contain data from 1961 to the present for all countries in the World. Once inside FAOSTAT at the menu page you can select the Domain of your interest and choose the Countries, Products, Elements and Years you need.

Q 2: Is there a limit in extracting FAOSTAT data? How many records can be loaded? Is that free of charge?
A. A non-subscriber is granted permission to view up to 500 data records; and is permitted to download only the first 25 data. If one wishes to access and download more data, this can be possible by subscribing to FAOSTAT On-line (please go to the bottom of http://apps.fao.org/ and click on Subscriptions for details).

Q 3: Is the FAOSTAT system available the whole day?
A: The system is not available for updating from 22.30 Sunday to 05.30 Monday Central European Time.

Q: Is it possible to export FAOSTAT data into Excel or other spreadsheets?
A: Yes, it is.- Once inside FAOSTAT domains,

1) Choose your query parameters,
2) Click on "Submit to Database",
3) Result will be shown,
4) Click on the "CSV file" option - This allows you to save in a CSV file to read and import in whatever spreadsheet you need, including Excel.

Q 5: With reference to the Country/Item list boxes for parameters selection inside each FAOSTAT Domain what are the meaning of "+" and ">" signs which appear before the Country/Item Aggregates names?
A: With reference to Countries or Items, the + sign indicates that the selection refers to Total figure for the Country/Item aggregate. The > sign indicates that the results will be shown by component of the selected aggregate (e.g. World+ will provide you with aggregate figures, i.e. World Total, World> will provide you data/figures for individual countries of the World).

Q 6: How is it possible to know the various definitions of Products and Elements in the FAOSTAT Domains?
A: Once you are in the Domain of your interest, you can see a blue line on the top of the screen. By double clicking on the word "Item" you will obtain a list of all items available in that Domain. Scrolling down the screen you can reach the product you are

interested in; by clicking on that line you will read a note about that product. The same is true for description of Elements.

Q 7: How is it possible to know the components of the Country/Item Aggregates in the FAOSTAT Domains?
A: Once in the Domain of interest, double click on the word Country/Item at the blue line on the top of the screen; a list of all Countries/items available in that Domain. Scroll down the screen to Country Aggregates an underlined name (e.g. AFRICA) and in the case of an Item Aggregate a name with "+" or ">" sign (e.g. CROPS PRIMARY>); by clicking on the line you will obtain the list of components and you can know which figures are included in the aggregate data.

Q 8: Why is it that when I am trying to extract a Food Balance Sheet all elements are automatically selected?
A: The Food Balance Sheet Domain is different from the other Domains as it generates a fixed Report. You can only select one country (to get "All" commodities for that country) or one commodity (to get all countries for that commodity). Elements are pre-selected. If you need individual selections you can refer either to Commodity Balance or Food Supply Domains which contain data corresponding to the FBS ones.

Q 9: How can I have methodological Notes on the FAOSTAT Domains?
A: Once in FAOSTAT, by clicking on the "Yellow icon" on the left side of each Domain line, it is possible to read methodological and statistical information referring to that Domain (including abbreviations used).

Q 10: What is the meaning of the following abbreviations found in some Domains of the FAOSTAT database: MT and Hg/Ha?
A: MT: stands for Metric Ton - Hg/Ha: stands for hectogram per hectare.

11) Q: I am a researcher working at ENEA (Italian Agency for Energy, Environment and New Technologies) in Rome. I would like to know how to get information on the FAO "data base on pesticide consumption".
A: Be informed that data on pesticide consumption are accessible on-line in FAOSTAT. Please note the Web address:
http://apps.fao.org/page/collections
subset=agriculture (Means of Production domain). Explanations about the database are available in the notes section of the "Means of production" domain at:
http://www.fao.org/waicent/faostat/agricult/meansprod-e.htm
or by double-clicking on the notes icon placed at left-hand side of the domain).

CONCLUSIONS OF THE CONFERENCE

As Mr. Luigi Biggeri has pointed out in his opening speech, "in a world of six billion people, the problem of providing food to everybody is one of the strategic themes of our society. At the same time, … the environmental and social impact involved in the transformation of agriculture has to be considered and studied".

At the same time, policies relevant for the agricultural sector have been broadened to take into account not only agriculture, but also rural development in a broader sense, reflecting the "multifunctionality" of agriculture. This concept refers to the fundamental links between sustainable agriculture, food safety, rural development, maintaining the landscape and the environment and (mainly for developing countries) food security. Therefore, traditional agricultural statistics need to be integrated with new statistics that take into account the changes that are taking place in the agricultural and food sector. It is also necessary to measure not only the quantity, but also the quality and safety of products, that have become increasingly important consumer concerns during the last decade. Moreover, there are public concerns about animal welfare issues, the preservation of biodiversity, agriculture's contribution to sustainable development, the protection of the environment and the preservation of the landscape.

In this framework, the role of statistics is fundamental to guide policy makers' and individual economic agents' decisions, at national and international levels. The international comparability of agricultural statistics in particular, has become crucially important in recent years, because of the globalisation of economic activities that make world economies increasingly interdependent. On the other hand, new technologies opened up new possibilities of conducting extensive surveys at a reasonable cost, as well as of disseminating results of statistical activities to the world community in a timely and easily accessible way.

The complexity of these issues is widely reflected in the very articulated programme of the Conference. Therefore, I do not have the ambition of summarising all the important conclusions achieved in the various sessions. However, thanks to the summaries prepared by all chairpersons, it is possible to identify some lessons learned during these days.

For example, one of the important issues discussed is represented by the fundamental question: "What is agriculture?" In fact, it is evident that in the past years radical changes took place in the agricultural and food sector, and the concept of "agriculture" has become more and more difficult to define. The most serious problems faced by agricultural statisticians include the difficulty in defining the boundary between agriculture and manufacturing and the choice of the observation unit, which depends also on the economic, social and institutional viewpoint.

Among the solutions proposed to solve these problems, the most promising seems to be the adoption of “flexible standards” which would allow the system of agricultural statistics to adapt to changing objectives. New international initiatives have been invoked to improve the dialogue among the various subjects and to fix new common definitions and classifications.

It was acknowledged that agriculture still plays a fundamental role in all societies, but with substantial differences. In developing countries, for instance, agriculture is the key for reducing chronic undernourishment and poverty, promoting economic growth and defending the environment. In countries in transition, agriculture plays a central role not only for reducing poverty, increasing growth and defending the environment, but also for managing social reforms. In developed countries, finally, agriculture has a key role for preserving the environment, promoting rural development and ensuring food safety.

Another issue that was widely debated was the role of new technologies, which in the course of the last decade have more developed and become widely available, changing the approach of statisticians to data collection, editing, management and dissemination dramatically. In the field of data collection, many innovative technologies have been developed, including remote sensing techniques and, more recently, the use of Internet for collecting data. These technologies allow collecting more information with higher statistical precision and, in many cases, at lower costs. The progressive reduction in the costs associated to these technologies make them more and more widely available and cost-effective. In developed and transition countries, moreover, administrative data are increasingly used for statistical purposes.

In data editing also, substantial progress has been made. However, in this area statisticians must be pay special attention to the quality of administrative data, since they are usually collected for non-statistical purposes.

Data dissemination is another field where important progress has been registered, thanks to the development of the Internet and to techniques like datawarehousing. However, statisticians, are facing new challenges, which include the development of new indicators to satisfy the emerging users’ needs.

The fundamental link between agriculture and environment was at the centre of the debate, and agri-environmental indicators have been discussed as fundamental tools for monitoring sustainable agriculture. The Conference encouraged national initiatives in establishing indicators to assess the environmental performance of agriculture in areas such as:

- protecting the stock of natural resources and landscapes impacted by agriculture;
- reducing environmental pollution resulting from agriculture;
- improving farm management and resource-use efficiency.

During the Conference it was felt that some areas need particular attention. From a technical point of view, more resources should be invested to provide assistance to countries in transition and to developing countries, and to disseminate existing and best practices widely. From the organisational point of view, the co-operation among international organisations should be improved and the possible transformation of CAESAR into a process should be discussed. Finally, from the political point of view, efforts should be made to improve the awareness of politicians about the fundamental role of agricultural statistics.

As already mentioned, these are only some of the conclusions achieved during the Conference. Now is up to national and international agencies to identify what they learned individually and as part of the international community in order to improve the quality (in the broad sense) of agricultural statistics and how to improve co-operation among them for meeting new users' needs and future challenges. The world is changing very fast and statisticians have the responsibility to help decision makers, at all levels, to find good solutions for the sake of this generation, as well as for that of future generations.

Enrico Giovannini
Chief Statistician of the OECD

ALPHABETICAL INDEX OF AUTHORS