Advancements in Survey Methodology and Analysis: Discussion

Francisco Javier Gallego

**Space Applications Institute/Agricultural Information Systems,
Joint Research Centre of the European Commission**

1. A General Comment: Innovation and Conservation

We can find in this session two types of papers:

- Presentations of large operational surveys, and
- Papers dealing with new techniques in agricultural statistics, at least in the sense that they have not been often applied to large operational activities.

It is interesting to find in the same session these two points of view that are the basic elements of consistent progress in agricultural statistics (as in many other fields of applied science and technology). We can often see resolute technicians claiming fast progress in front of conservative administrators who do not trust very much approaches that they judge adventurous, and possibly jeopardizing the success of operations. They generate a sort of dialectics that puts thesis and antithesis in front of each other to produce a positive synthesis that is the engine of improvement.

Prudent officers (sometimes seen as fearful) can play a very positive role in the process of development consolidation, as long as they give the opportunity to test new procedures. Their incredulity pushes technicians to make more robust proceedings and reduces risks of later large scale rejection.

Remote sensing in agriculture might be a good example of a powerful tool that has suffered the consequences of an excessive optimism in its early days Terms such as “spectral signature of a crop” indicate the expectation that precise agricultural estimates might be soon produced without ground observations. The reality turned out to be more modest, and the credibility of remote sensing went through a dark period. Maybe one of the reasons for this phenomenon was that the initial enthusiasm did not have a sufficiently serious skeptical counterpart.

2. “The National Resources Inventory and its Role in U.S. Agriculture” by Jeffrey Goebel, Natural Resources Conservation Service, USDA

The National Resources Inventory conducted by the Natural Resources Conservation Service is certainly a very impressive operational survey that witnesses a long history of care on natural resources, and particularly on soil conservation, that has later become a mayor topic of concern everywhere. The survey approach combines segment and point sampling techniques, photo-interpretation, land cover observation on the ground and more complex physio-chemical measurements.

From the point of view of sampling design, it can be considered a worldwide reference with a complex stratification and variable size of segments to improve the efficiency. Data from the survey are made available to a wide variety of users, which is an important element to justify the very important human and technical resources invested in this survey.

3. "Land Cover and Land Use Classification using TER-UTI" by Christian Gay & Jean-Claude Porchier, Service Central des Enquêtes et Études Statistiques, France

This paper presents a relatively conservative way of tackling area estimation that does not produce huge revolutions, but provides a simple, cost effective, quite convenient for adaptation or direct application in developing countries.

The long experience by SCEES at running this survey has lead to a safe sample design and balanced set of instructions for ground enumerators. The system provides a basis to subsample for surveys focused on a particular topic, and the need of making the difference between land use and land cover is also taken into account in the nomenclature.

Considerable improvements can be made in the sample design. Partial sample rotation would improve the precision of estimates in a particular year using the information of previous years. If some proxy information on the territory is available in a GIS environment, the sampling plan or the estimation procedure may be improved, in particular in "hot spots" such as peri-urban areas or sensitive areas. Most possible improvements require integrating these rich data in a GIS environment by a precise location of points. This concern now seems to be emerging.

The supervision of 1/16 of the sample certainly gives interesting information — confusion matrices and a basis to discuss the problem of distinguishing between location errors, identification errors, or nomenclature interpretation problems. Unfortunately this information is not provided in the paper.

Comparison of National Resource Inventory (USDA) and TER-UTI

	NRI	TER-UTI
sample size	300,000 segments (PSU) ~3 points/segment 800,000 points	~15,500 segments (PSU) 36 points/segment ~550,000 points
frequency	every 5 years	every year
observation land cover	80% photo-interpreted 20% ground visited	100% ground visited
items observed	Land cover / land use complex measures in visited points	Land cover / land use
stratification	complex stratification	no stratification
PSU shape/size	square / variable size	square / constant size

4. "Utilization of Document Imaging Technology by the 1996 Canadian Census of Agriculture" by Mel Jones & Ivan Green, Statistics Canada

We find here a nice example of enthusiastic people implementing a new technology in a large operational activity, and their (presumably friendly) fighting with skeptical staff who would have rather selected a more traditional approach.

This paper is quite convincing about the fact that imaging technology will soon substitute paper for large-scale archiving and retrieval, in the same way that many of us increasingly prefer storing office

documents as files in a PC than as paper. This way of organizing your office with less paper can be safe if a good backup infrastructure is available, such as an automatic daily backup of some directories, but may become dangerous otherwise.

In a similar way, some careful assessment is needed before exporting to other countries the Canadian example of introducing imaging technology for the agricultural census, carried out under particular conditions: relatively low number of farms, highly efficient staff, and, as the paper points out, some luck. The guarantee of document integrity seems insufficiently clear in the paper, and cost comparison with traditional methods should also be evaluated after having the point of view of more conservative people.

5. "New Methods of Editing and Imputation"
by Svein Nordbotten, University of Bergen, Norway

This paper can also be classified into the category "enthusiasm for new methods". Although artificial neural networks (ANN), whose application is the main topic of the paper, are not at all new as a concept, their development has been possible only when powerful computers became common, since ANN generally require a high amount of computation.

The paper discusses the application of ANN to statistical editing (outlier detection) and imputation, both to cope with non-response and to produce small area statistics with the help of ancillary information. The range of topics covered by the paper is quite wide, and consequently specifics cannot be given, but the comparisons with other methods look rather convincing.

ANN provide an alternative to more classical techniques in many fields. Outlier detection methods have been available for a relatively long time [Barnett and Lewis 1978, Hawkins 1980], and a number of imputation procedures are available in the literature [Särndal et al. 1992, Chapter 15]. Performance comparison between ANN and more classical methods is not always made and when carried out, it is often done by ANN specialists with little experience in classical methods. In many papers, the reader may think that contests should be organised to get a more clear assessment of performance and robustness. This paper is more convincing than the average of ANN papers about the fact that comparison with other methods has been made under conditions of fair competition.

6. "Partial Overlap Samples and Combining Information"
by Wayne Fuller, Iowa University

This paper discusses the advantages and limitations of partial rotation of samples for multitemporal surveys. The consideration of eigenvalues and the trace of the covariance matrix for the vector of targeted estimates is suggested as a trade-off criterion to select between a fixed panel or a supplemented panel (partial rotation). A method is also briefly presented to improve consistency in multipurpose estimates.

Maybe some kind of weighting can be introduced to take into account the importance that users give to the estimation of a variate at a certain time compared with the importance of the estimation of differences. For surveys involving personal interviews, sample rotation also has a positive effect on the response quality. For land cover surveys, the problem is different — the cost of partially renewing the sample may be important and must be considered to determine a sample rotation strategy.

The topics and methods discussed are without any doubt of interest and promise possible application in many countries, but communication seems to be a major obstacle in many cases. In spite of the effort made by Prof. Fuller to simplify the presentation in this paper, introducing, for example, an assumption of an autoregressive process with known correlation, the text still remains difficult to read for many official statisticians, and the availability of user friendly software to carry out generalised least squares estimators is still a question mark for many institutions. In the example of the TER-UTI survey presented in this session, people are used to an estimation operation that is as simple as multiplying the number of points on a certain land cover by 1 km^2, and there will be some resistance to substitute this operation by something involving the inversion of large matrices. Presenting generalised least squares in an attractive way to official statisticians is a major challenge for mathematical statisticians.

7. "Small Area Estimation in India - Crop Yield and Acreage Statistics"
by B.D. Tikkiwal, Institute of Developmental Research and Statistics, India
& G.C. Tikkiwal, J.N.V. University, Jodhpur, India

This paper gives an interesting description of some sampling and estimation designs for agricultural statistics in India, although the reader that is not familiar with the Indian administrative structure may have some difficulties to follow without making a scheme of the meaning of AAO, VEW, chaks, mandal farmers, etc.

The main contribution is a comparison of composite estimators that combine synthetic estimators with generalized ratio estimators. The approach might overcome some limitations of synthetic estimators and looks easier to apply than other small area estimation procedures that have been used in agricultural statistics [for example, Battese et al. 1988]. Some additional clarifications would be of interest on the computation of variance from a single sample. If the results presented are confirmed in other countries, the method would be of interest, and not only for developing countries, as stated in the paper. Actually, India is a developed country if we speak about statistics.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, 83, pp. 28-36.

Barnett, V. and Lewis, T. (1978), *Outliers in Statistical Data*, New York: John Wiley & Sons.

Hawkins, D.M. (1980), *Identification of Outliers*, Chapman and Hall.

Särndal, C.E., Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York & Berlin: Springer Verlag.