January 2014

E

منظمة الأغذية والزراعة للأمم المتحدة
联合国粮食及农业组织
Food and Agriculture Organization of the United Nations
Organisation des Nations Unies pour l'alimentation et l'agriculture
Продовольственная и сельскохозяйственная организация Объединенных Наций
Organización de las Naciones Unidas para la Alimentación y la Agricultura

# Asia and Pacific Commission on Agricultural Statistics

| Twenty-fifth Session |
| --- |
| Vientiane, Lao PDR, 18-21 February 2014 |
| Agenda Item 8 |
| A new approach to food balance sheet methods and design |

**Contributed by:** *FAO Statistics Division*

## 1. Overview

The compilation of food balance sheets (FBS)[1] has a long history, extending as far back as 1934. The importance of FBS is demonstrated by their wide use by aid and development agencies, by public health officials, and by research economists and other scholars. Perhaps the most visible use of FBS data is in estimating levels of food and nutrition insecurity and in the monitoring of food security trends over time. In this context, they are used to estimate the prevalence of undernourishment, which in turn measures progress towards the Millennium Development Goal (MDG) 1.9.

Since the methodological conception of FBS, substantial changes have occurred within the agricultural sectors and in information and communications technology (ICT). The agricultural sectors saw changing food demands by consumers; changing and new found demands for crops by the non-food sector, including livestock feeding and feedstock for industry and energy;

[1] FBS provide a comprehensive picture of the pattern of a country's food supply during a specified reference period. Currently, they show the sources of supply and utilization for each food item, that is, each primary commodity and a number of processed commodities potentially available for human consumption. The *per caput* supply of each such food item available for human consumption is 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, in terms of caloric value, protein, and fat content.

globalization and global value chains, made possible by technological breakthroughs in transportation and ICT; and new and improved agricultural technologies that transformed primary foodstuffs.[2] Growth in ICT has also expanded and automated many tools and processes available for data collection, processing, verification, compilation and dissemination.

These changes necessitate an overhaul to both the methodology underlying the FBS, as well as the methods and workflows behind them. This is further driven by the demand for more reliable, credible and timely evidence on food and nutrition security.[3] Furthermore, the current human resource-intensive approach, combined with ICT growth, provides an opportunity to exploit technological innovations and reallocate human resources to higher value-added activities, such as research into country food situations and the changing dynamics of the agricultural sector.

The proposals presented in this paper build on a review of the FBS, and include new approaches to its design, measurement and classification. In parallel, a new statistical working system is being developed to automate data collection, processing and compilation.

The philosophy of the new approach is predicated on the following elements:

- **Empirically data-driven Commodity Accounts**. The new approach recognizes that much of the data required to populate the current FBS framework remain unobservable or, at best, sparse. To respect data availability and preserve information, no attempt is made to construct accounts for commodities with insufficient or missing data.
- **Plausibility**. For those commodities for which accounts are created, commodity usage is imputed in countries where it is not observed. Using empirical data on types of utilization, hypothesized determinants are tested and, if valid, used for imputation. When new data arrives, models are re-estimated and updated to improve their predictive performance.
- **Exploiting all available information** about the phenomenon in question to improve estimates is a core principle of the new approach. This principle is applied in imputing missing production and trade data, and also in estimating commodity utilization as in the case above.
- **Exploiting technology**. For processes that do not require human intervention or when such interventions can be automated, algorithms and appropriate script are employed.
- **Rules-based**. FBS are balanced according to rules governing the physiological, economic and historical bounds of utilization, as well as “priors” that describe our certainty about the levels of utilization.
- **Timeliness**. The rapid compilation of FBS with a minimal time lag is an overarching prerequisite, meeting demands for timelier monitoring on food security.

[2] Knudsen, M. T. et al (2005). *Global trends in agriculture and food systems*. **Global Development of Organic Agriculture: Challenges and Promises** (Eds N. Halberg et. al.). Cambridge: CABI Publishing.

[3] The current framework for FBS imposes stringent and often unrealistic demands on obtaining data beyond primary production and basic trade flows. Furthermore, the empirical basis for imputing missing data is often limited our outdated, often using historical conversion factors and share parameters, with corresponding metadata often absent.

The paper is structured as follows. Section 2 describes innovations in the building blocks and design of FBS. The remainder of the paper then describes innovations to consistent imputation of commodity utilization. Finally, it provides recommendations to improve FBS quality and reliability.

## 2. FBS foundations

### 2.1 Core data

Production and trade data are important elements of the current FBS production system, and will continue to remain central to any new system. Consequently, data collection efforts, both in the production and trade domains, remain key. Since the quality of FBS can ultimately only be as good as the core underlying data, concerted action is required to improve data inflows. In the meantime, there is substantial scope to exploit innovations in ICT in data collection (or “data harvesting”), data processing and data management.

Production data will continue to rely on country-level questionnaires, where currently problematic response rates of 60% are compounded by incomplete questionnaires. This reflects, in part, limited national resources and capacity in measuring productive activities, and insufficient investment in statistical infrastructure. Internationally coordinated efforts, such as the Global Strategy to Improve Agricultural and Rural Statistics, are expected to provide some remedy, albeit not in the short-run.

In the short-run, the new approach will exploit all reliable data from external sources, including web-based data, data published in electronic reports, data disseminated in other databases, and data from information systems related to commodity activity and food security. Remaining data sparseness will be addressed through imputation methods based on country specific information and historical regional trends, while accounting for the correlation between country and regional fluctuations, and changes in climatic conditions, such as weather shocks, which affect year-to-year production. Under the new approach, imputed values have been shown to display stable characteristics, while taking into account environmental conditions, and maintaining the ability to incorporate further information on the determinants of production outcomes, such as precipitation, temperature and vegetation.

Trade data will come from the United Nations Statistical Division’s (UNSD) Monthly Comtrade database and its new web-based application for the dissemination of monthly merchandise trade statistics. Relevant trade data is extracted using an application programming interface (API) software. The shift towards using Comtrade, rather than processing country-level electronic trade files within FAO, exploits the trend among international and regional organizations that sees greater cooperation in statistical areas of work. This cooperation has resulted in more specialization in areas of comparative advantage, reduced duplication in statistical data development, and improvements of statistical products and databases to meet the needs of multiple users. It has also reduced response burden on countries in reporting the same data to multiple organizations, and the burden on FAO to process trade files and apply appropriate commodity classifications. Under the new design, data are extracted, validated, imputed and consolidated, using a set of orderly rules.

### 2.2 Classifications[4]

FBS commodities are currently classified according to the FAOSTAT Commodity List (FCL). The FCL was developed in the ‘60s, based on the UN Standard International Trade Classification (SITC), and its structure reflects “commodity trees”, where primary crops and livestock constitute “parent” items and their derived products are the “child” items.

Originally, the aim of the FCL was to provide a common classification framework for production and trade statistics on crops, livestock and their derived products in order to construct supply and utilization accounts. However, this involves substantial and often arbitrary conversion of trade data, provided by countries in Harmonized System (HS) format, to FCL. In addition, the HS has evolved over the years as new versions have been developed, while the FCL has remained a “static” tool as it has never undergone periodic review.

The UN Central Product Classifications (CPC) was developed in 1998 building on the HS and ensuring harmonization with the UN International Standard Industrial Classification of All Economic Activities (ISIC). The need for the development of the CPC originated from initiatives in the early 1970s to harmonize international classifications, based on the general agreement on the need to improve harmonization among the various classifications prepared under the auspices of the UN and other international bodies.[5] However, the structure and list of items in the first versions of the CPC lacked the necessary detail to be relevant for agricultural statistics, until CPC Ver.2 was endorsed at the 37$^{th}$ United Nations Statistical Commission in 2006.

CPC Ver.2, as well as HS 2012, significantly increased the detail on agricultural products. They were the result of the continued collaboration of FAO with other International Organizations and with member countries to improve international product classifications relevant for the

[4] The primary purpose of statistical classifications is to promote consistent and comparable information by providing a framework with internationally recognized definitions and concepts, and groupings. Statistical classifications are used to harmonize and coordinate data collection, compilation and dissemination; aggregate and disaggregate data sets in a meaningful way; support policy and decision-making; and enable inter-group, inter-regional and inter-temporal analyses and comparisons. International classifications provide this framework at the international level to enable inter-country comparisons.

[5] http://unstats.un.org/unsd/cr/registry/regdntransfer.asp?f=145

agricultural sector. The recently endorsed CPC Ver.2.1 and the development of an expansion for agriculture, developed and implemented by FAO (called CPC "expanded"), will provide even more detailed information on agricultural products, and further increase the degree of harmonization between CPC and HS.

These improvements to the CPC and HS set the conditions for FAO to adopt these standards as reference classifications for statistics on agricultural production and trade, in place of the FCL. The greater granularity of the CPC and HS enable them to mimic the FCL without loss of accuracy, while keeping meaningful aggregations and providing improved definitions and codes. Furthermore, trade data can now be standardized to the CPC for the compilation of FBS, reducing the workload associated with conversion to FCL. Moreover, adoption of the CPC is expected to reduce reporting burden by countries, increase data comparability across countries, and facilitate the integration of agriculture into national statistical systems (in line with the second pillar of the Global Strategy), enhance compliance of FAO's standards with international ones, and improve the transparency and governance mechanisms for the management and review of classification systems in the Organization.

### 2.3 Which commodities?

An inspection of the cumulative percentage of Dietary Energy Supplies (DES) contribution by commodities reveals that about 60 primary commodities account for 99 percent of total caloric intake at the global level. This suggests significant scope for reducing the list of commodities currently managed by FBS, which could increase efficiency and improve timeliness with potentially minimal impact on its uses or users.

A second proposal would be to restrict food accounts to primary commodities, such as wheat, rice, and maize, and to remove commodity accounts for processed foods, such as flour, vegetable oils, sugar, beer and food preparations. Production data are sparse for the processed food commodity accounts, which cover over 600 items. Focusing commodity accounts on primary commodities only would restrict commodity standardization to the trade domain, where standardization, or "pre-standardization," reduces derived or processed products to their parent commodities. This would address a major weakness of the current system in terms of the intractability of the current standardization method. The move, nonetheless, still requires accurate technical conversion factors. Concerted efforts are underway to collect these coefficients.

## 3. New methods to estimate utilization
### 3.1 Livestock feed demand

The need for a new approach to estimate utilization of livestock feed in the current FBS framework stems from a host of issues: (i) in many countries it is estimated as a residual in the balance of demand and supply; (ii) where official estimates are missing, the approach to calculate feed use so far has been "supply-driven", in that it considers feed as a given percentage of a commodity available in the country, regardless of the potential demand for feed in the country as dictated by the livestock sector, as well as the efficiency of feed; (iii) by implication, changes in livestock populations and herd structures, production intensification, losses, mortality, etc. are not captured by the current framework.

The rapid intensification of livestock sectors deserves further elaboration. In recent decades meat and dairy production has been increasingly commercialized globally, with emphasis on large scale, "factory-farming" type production systems. Some of the drivers and impacts of this phenomenon are listed below:

- Increased demand for livestock products.
- Economies of scale, as commercialized production of livestock and livestock products lead to lower per unit cost of production.
- Efficiency gains resulting from rapid strides in the livestock production technology, and ease of transfer of new technology facilitated by globalization, have led to efficiency gains.
- Changing structure of markets arising from more vertical integration in the food sector, which has led to mass production and an emphasis on consistency of quality, traceability and food safety.
- Environmental concerns, which have encouraged better conversion of feed, leading to lower harmful emissions such as N2O, methane, etc.
- Increasing population pressure (or, decreasing land to population ratio), resulting in increased demand for both crops for food (competing resource for feed) and crops for feed (due to more demand for livestock products) prompting the need to make better utilization of available resources for feed.

***A demand-driven approach***

The basis of the demand-driven approach for feedstock estimation is the energy and protein requirements of a herd. These are the two building blocks of nutrients for animals. Feed grains and other high energy feeds generally are combined with oilcakes or other high protein meals, as necessary, to achieve the required energy and protein content and other requirements of feeding systems. In more commercialized or intensive livestock production systems, desired nutrients are provided mainly through such feed concentrates. The relatively rapid increase in the feed use of concentrates can be linked to major changes in the composition of the livestock populations and outputs. In addition, shifts in the composition of livestock output toward products from monogastric animals have favoured the increase of concentrates.

The proposed method distinguishes two steps:

*(i) Determining feed requirements for metabolizable energy and protein*

The basic rationale for changing the estimation method for feed use is to align it more with the feed requirements and then allocate needs to available grain and protein feeds. The method thus distinguishes two steps: in the first step, total requirements in terms of Mega Joule Metabolizable Energy (MJ ME) and digestible protein (DP) are determined; in the second step, the requirements are mapped back into available concentrate and compound feedstuffs, i.e. into oilcakes, oilmeals and cereals.

For the first step, feed needs are expressed in terms of Grain Consuming Animal Units[6] (GCAU) and High Protein Consuming Animal Units (HPCAU). The GCAU is defined as a weighted aggregate of the number of all livestock types produced in a country, with weights (W) adjusted over time to reflect changes in yields (Y), such as carcass weights or milk per milking cow. The weights, which represent a combination of biological factors and feeding practices, convert and express all livestock types into poultry equivalents. They are adjusted for livestock-specific intensity rates and livestock-specific efficiency rates. Just like yields, intensity rates and efficiency rates change over time.

In practice, changes in intensity and efficiency rates are often off-setting over time, i.e. increases in feeding intensity are often associated with parallel improvements in feeding efficiency. In a final step, the CGAU is then expressed in terms of MJ ME, taking into account the energy needs for poultry. The same logic is applied to calculate total protein requirements, creating a HPCAU in parallel to a GCAU. There are no explicit calculations for non-compound feed. At intensity rates of unity, use and requirements are identical.

The overall framework is applied to six types of livestock, with one module each for cattle, pigs, poultry, small ruminants (sheep and goats), aquaculture and others (mainly, horses and camels). The structure of the model and the equations are mostly the same for all livestock modules, and variations are with respect to assumptions for the parameter values. The structure of the module used in the case of aquaculture was slightly modified in that the feed conversion ratio is employed as the basis for calculating feed as opposed to energy requirements. Work is also underway on capturing changes in feed efficiency (the amount of compound feed needed to produce a given unit(s) of meat, milk and eggs) as well as changes in intensification.

[6] The United States Department of Agriculture (USDA) also employs the term "GCAU", however, it differs conceptually from the FAO approach in that their GCAU (in index form) is used to verify estimated changes in feed use, which is treated as a residual in accounts.

*(ii) Determining allocation of compound/concentrate feedstuffs to match requirements*

Once total requirements are calculated, the allocation of individual feedstuffs follows in a second step. Again, the allocation is undertaken into distinct steps. In a first step, all available oilcakes and oilmeals are mapped against requirements. All available oilcakes and oilmeals are, with few and minor exceptions, assumed to be used for feeding purposes. In a second step, after all available oilcakes are allocated to the different livestock, the remaining protein requirements are calculated and assumed to be covered by other compound and concentrate feeds, which come mainly, though not exclusively, from cereals. The allocation across the various cereals (maize, barley, wheat, etc.) is based on prior information of typical, country-specific feeding practices.

## 3.2 Food losses

According to the definition of FAO, food waste is any edible material intended for human consumption, arising at any point in the Food Supply Chain (FSC) that is instead discarded, lost, degraded or consumed by pests.[7] In the accounts of FBS, food losses are considered to be the amount of lost commodities, starting from the moment when production is recorded until it reaches the consumer. Food wasted at the household level is not considered as a loss in the FBS, but rather as “waste”.

The FBS reports food losses for each relevant type of food, each country and for each year. However, the amount of food loss is not asked in the country’s production questionnaire. The calculation of food losses are usually based on some fixed ratios, though many ratios have not been updated since the 1960s.

*Amount of food loss = (Total production + Imports + stock variation) x Loss ratio.*

***Updating loss parameters***

Unfortunately, empirical data on food losses are rare. The ideal is properly measured data of food losses for each commodity by country and over every year, but this is far from feasible. Within the FBS, the share of ‘measured’ data, which are based on loss assessments from national institutions, is smaller than 5 percent. The remaining over 95 percent of data is out-of-date and requires updating.

Using ‘measured’ FBS data, together with data from additional surveys, we impute loss ratios for cases in which no measured data are available. In the first step of this imputation

[7] FAO (1981). *Food loss prevention in perishable crops*. FAO Agricultural Service Bulletin, no. 43. FAO Statistics Division: Rome, 1981.

procedure, we select *a priori* those factors that conceivably influence food losses. For example, countries' infrastructure might play an important role for losses during storage and transportation. Inadequate storage facilities should raise the ratio of losses. Similarly the quality of roads in a country is supposed to influence losses during transportation. Other important factors include climate zone, a country's share of small farms, rainfall at harvest, presence of pests, etc. An important non-linearity arises in the case of "excess production". That is to say, when bumper harvests are obtained, a country may not have the infrastructure to store the excess, or the price incentive to distribute to markets. In such cases, losses could be greater, especially for perishables.

After selecting the principal factors, which tend to be different between commodities, we estimate the impact of these factors on national loss ratios with a regression model. The parameters of the regression estimations are then used to predict national loss ratios for any given commodity and countries' characteristics. In this way, loss ratios that are out of date can be easily updated. In order to guarantee reliability of the results, all imputed ratios will be reviewed by experts for their plausibility.

### 3.3 Stocks

Estimating stocks in FBS is difficult task, given that total stocks are rarely measured or observed, and tend to be estimated indirectly as residuals from the balance of domestic utilization and domestic supply. This raises, however, important issues. First, stock estimates inherit the measurement errors of all other variables when estimated as a balancing residual. Second, as stocks include, by definition, amounts carried over from the preceding year, they accumulate errors from all previous years' estimates. This is problematic when errors do not average out and stock levels are of interest, and not just their variations. In other words, errors made in 1970 may still be present in the latest stock figure, as the stock estimate possesses "infinite memory" of past errors.

This led to work on model-based stock prediction built on the theory that stocks are a function of endogenous prices, and hence, can be determined from prices alone. The model has the drawback that robust inference requires large samples of more than the 20 annual price data points currently available. As a result, a simplified model was employed to infer stocks from prices using pooled data from the CCBS database of the FAO Trade and Markets Division. However, as well as a general lack of statistical significance, the counterintuitive result of coefficients having the wrong signs was often obtained. It is expected that initiatives to improve data collection on stocks, such as the Global Strategy, will help address this issue.

### 3.4 Food

The basic idea of the new approach rests on employing all available information to guide estimates of utilization. Food utilization is no exception to this rule. There is a realization that problems often arise with regard to the coverage/quality of the basic data that drive domestic supplies. These centre on the potential to underestimate calories available for intake in that some food commodities are excluded from national production statistics, and by implication are not captured by FBS. For instance:

- An appreciable part of total production may be non-commercial or for subsistence, i.e., foodstuffs grown or gathered wild by households for their own consumption.
- There are crops that are continuously harvested at irregular intervals over a long period of time, such as cassava, and certain fruits and vegetables.
- Many food crops are cultivated using multi-cropping techniques.
- Significant amounts of trade across national boundaries go unrecorded.

Another set of problems relates to missing data that are imputed. It stands to reason that when basic data are incomplete and unreliable, an estimate of food available for human consumption may be inaccurate. Furthermore, as food utilization is often derived as a residual in the balance, the error is unquantifiable and its direction is also unknown. For these reasons, a solution would be to independently corroborate and correct the estimate of this component from other existing statistical sources of information. One such source would be from Nationally representative Household Surveys (NHS).

The use of survey data can reduce reliance on the residual or balancing approach in arriving at food availability estimates, while providing more flexibility in handling the other elements for which basic statistics are poor. NHS can also serve to improve the FBS estimates with respect to the underreporting of commodities. There are additional advantages, which include: broad consistency checks on consumption patterns (contribution of various food groups in total intake); information on levels of waste (food available versus food consumed); and assessments about the variation of food availability/intake among households. These surveys cannot, however, provide a surrogate for FBS due to lack of frequency and incomplete coverage, among other issues.

A methodological framework for incorporating information from NHS into FBS, with the ultimate aim of enhancing the accuracy of food measurement is being developed. The problem is not new, as the similar task of reconciling survey data with national accounts is well recognized in the literature, and guidance from prior research will be sought.

### 3.5 Industrial use

Agriculture plays an increasingly important role in the industrial economy, providing feedstocks for the production of liquid fuels, chemicals and advanced materials, such as composites for industry, while the emergence of green industries expands opportunities for the rural sector. Furthermore, biological science has the ability to make both incremental efficiency improvements, and to bring about wide change in a host of farming sectors through use of enzymes, fermentation and organisms for processes and products in the energy, chemical, pharmaceutical, food, textile, and pulp and paper industries. Much of this potential is already being realized, especially when considering the rapid growth of the bio-fuel sector: ethanol is produced from easily fermentable agricultural feedstocks such as sugar cane, sugar beet, cereal grains and cassava; bio-diesel is produced from vegetable oil (typically rapeseed, soybean and palm oil) using a process of chemical modification; and the starch industry extracts and processes starch from cereals, roots and tubers for use in products, ingredients and functional supplements in food, feed and non-food applications

A generalized model framework is required to provide reliable estimates of industrial usage. Important drivers of industrial usage likely include income, the level of industrialization, policy mandates, and a country's comparative advantage in cultivating and processing the crop. To test and refine the model, a data collection strategy is necessary. This can build on data collection strategies that inform other important industrial activities involving food crops, such as the Aglink-Cosimo framework of the OECD/FAO, which provides data on bio-energy production by feedstock, and the USDA data collection on starch production.

### 3.6 Seed

Among the different categories of utilization, seed use is an ideal candidate for modelling, *a priori*, according to a deterministic rule. Seeding rates and, ultimately, the demand for seed, can be modelled as a function of target plant density, establishment percentage and seed weight. Multiplied by area planted, seed use can then be derived. Of course stochastic factors, such as agro-climatic differentials, seed size, germination, timing, and varieties, can lead to variation in seed use. Knowledge of these variables also allows modelling of seed retention based on area harvested.

However, with the rise in high-precision commercial farming, many farmers choose to use certified seed purchased from specialized seed farmers. This trend requires the need to capture commercial seed production quantities, say through the annual production questionnaire, as well as adjusting for seed retention rates.

## 4. Putting it all together

The current practice of balancing FBS does not well incorporate knowledge of a country's agricultural and food situation, largely due to the lack of resources to research how domestic supplies are precisely utilized in a country. The primary aim in balancing, which is left to the discretion of the statistical clerk, can become the removal of statistical discrepancies, at the expense of introducing implausible adjustments to utilization.

In this regard, recognizing that an existing wealth of information can be introduced, *a priori*, via computational methods, will enhance the ease, reliability, precision and coherence in determining utilization. Substantial efforts have already been directed towards collating this knowledge, while techniques such as Bayesian estimation are well established in the literature for balancing underdetermined systems.

These methods will never replace the need to judiciously review the quality and credence of FBS, but they can serve as a very useful aid towards this task. Indeed, a rules-based approach to balance FBS, informed by theory, evidence and our certainty about prior information, will alleviate the laborious and costly task of balancing by discretion, also improving on plausibility. More importantly, the rules-based approach will also free-up considerable resources to improve the underlying data of FBS, allowing human resources to add value where it is most needed.

## 5. Recommendations

This paper highlighted new and on-going innovations in FBS design, measurement and classifications. Towards the broader goal of improving the quality and reliability of FBS, APCAS member countries are requested to discuss and consider the following recommendations:

1. Member countries improve the collection and dissemination of timely and reliable data on food supplies that conform to the latest version available of international classification schemes, notably, CPC for production and HS for trade.
2. Member countries produce their own FBS, providing reliable and detailed information on food utilization, while FAO FBS activities focus on: (a) assisting in the FBS compilation process; (b) adopting the most recent methodological improvements; *and* (c) helping create necessary statistical capacity at country level.