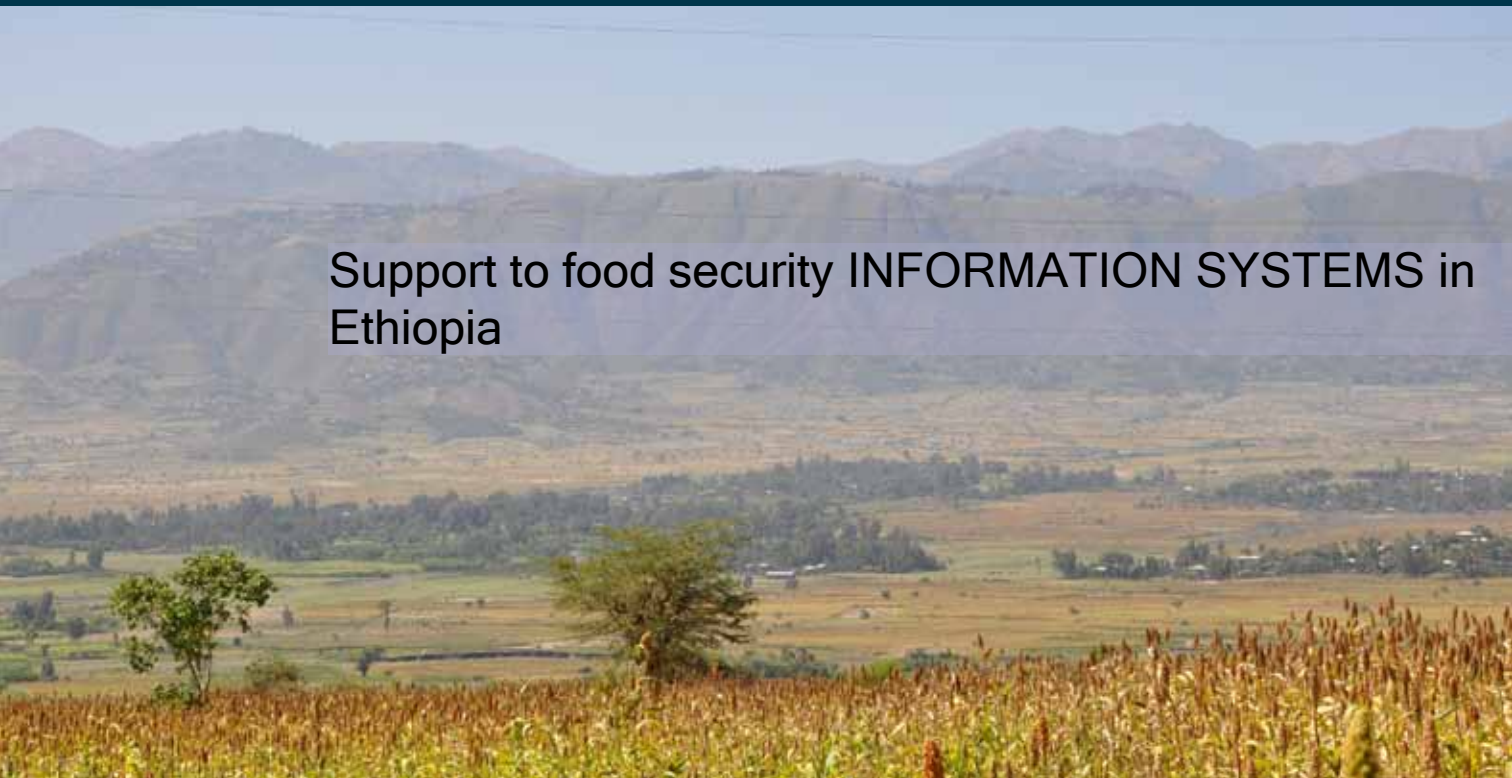
Harmonizing Agriculture Statistic Programs
Small Area Estimators

Project Funded by the European Commission
CRIS120-352
GCP/ETH/071/EC

Ministry of Agriculture and Rural Development

National Meteorological Agency



Support to food security INFORMATION SYSTEMS in Ethiopia

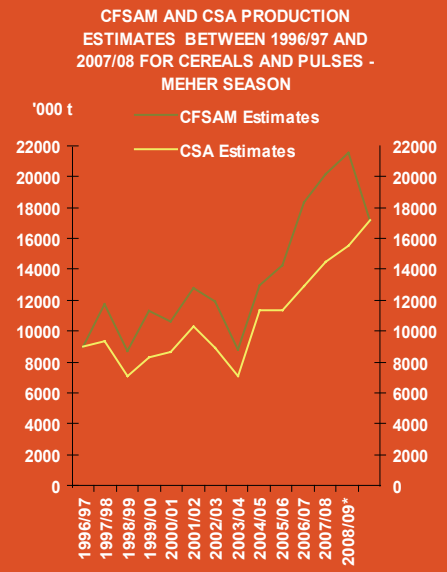
‘Support to Food Security Information Systems in Ethiopia’ is a project valued at 2 million euros, funded by the European Commission and implemented by the Food and Agriculture Organization of the United Nations (FAO). The project works closely with the Government of Ethiopia (GoE) and in particular the Central Statistical Agency (CSA), Ministry of Agriculture and Rural Development (MoARD) and National Meteorological Agency (NMA). 1.Background.Estimates of crop production are among the most important information required by national institutions. Before the project, crop area and yield estimates were based on field surveys carried out independently by CSA and MOARD with significantly different results. Diverging estimates had serious implications on decisions concerning agricultural strategies, policies and food aid allocations and distributions (see figures 1 & 2). For this reason, a better estimation of agriculture production is considered a national priority. A suitable methodology building upon the strengths of both CSA and MoARD systems has been defined with assistance from the project. While CSA implements a statistically sound methodology to generate mandated estimates at Federal, Regional and Zonal levels, MoARD generate estimates at the critical Woreda level.The project is assisting in the implementation of a robust statistical methodology assisted by the latest technology, to develop and enhance the complementarities of the two systems. At the core of this statistical methodology, CSA and MoARD datasets are integrated in one database reporting up to Woreda level using the Small Area Estimates (SAE) statistical technique.

2. Project Component Objective.The overall objective of the project is to define a suitable methodology to improve the reliability of agriculture statistics. The project is introducing cutting edge technology, which includes amongst other things, advanced statistical tools such as Small Area Estimators.

3. CSA Data Generation System.CSA conducts “probability agriculture surveys” whereby estimates of crop production are made based on a sample of plots measured. Probability surveys require a sampling frame. The units listed in a sampling frame are assigned probabilities that are then used for selection of the sample if a survey is undertaken. The current master sampling frame comes from the data obtained from the population and housing census and used for the surveys under the Rural Integrated Household Survey Program (RIHSP). The sample design of the RIHSP is a stratified, two-stage, systematic sample, with administrative zones as strata. Enumeration Areas (EAs) are territories smaller than a village with a known number of households. These are used as primary sampling units (PSUs). The sampling frame includes over 90 000 geo-referenced EAs in Ethiopia. EAs are selected using systematic selection procedures where probabilities are proportional to the number of rural households listed in the 2007 population census. Enumerators visit the selected EAs and list all rural households living within the EA for every survey. Once the updated list of rural households is obtained, a systematic sample of 30 farming households from the new listing is selected for the second stage sampling.

From the master sample, 2 200 EAs are selected for the main Meher agriculture survey, households are listed in each EA and 30 farming households are then selected systematically from the new listing for data collection. For each of the 66 000 selected households, farmers are asked how many fields they have and what crops they are growing in their fields. The fields are then visited and measured with meticulous care to obtain the hectares in each field and yield of each crop. Nearly 600 000 fields are measured every Meher season and on one third of them, crop cutting is conducted to estimate yield. In addition, field level economic data are collected through interviews with farmers.

Past divergence in crop cultivated area and production estimates between CSA and the FAO/WFP CFSAM estimates - based on MoARD data - had stabilized in the range of 2 million hectares whereas the difference in production had stabilized at 5.5 million tons since the start of the project two years ago. Figures 1 and 2 show that in 2008/09, the successful definition of a suitable methodology for estimating agriculture production has eliminated the divergence and thus consolidated agriculture statistical program in the country.



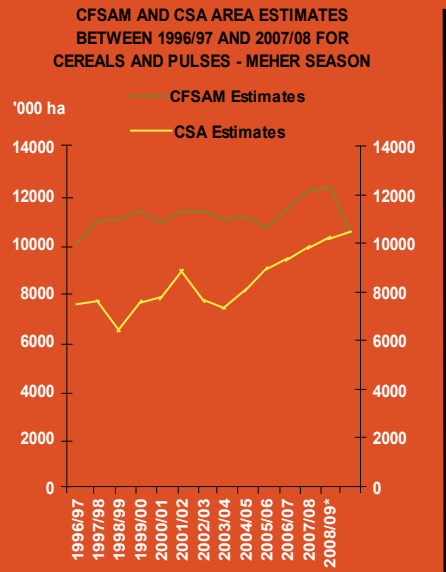


Figure 1 and 2: CFSAM and CSA production estimates between 1996/97 and 2007/08 for cereals and pulses - Meher Season.

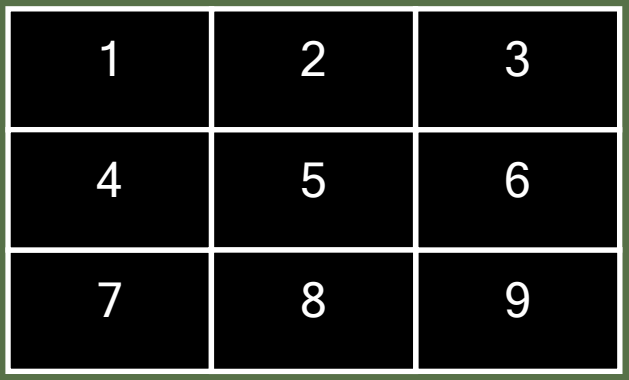
From a sample of 66 000 farm households, an estimate is calculated of the number of hectares and production of teff, wheat, maize and all crops that are currently planted as well as the farm economic data using the proper “raising factors.” (the term “raising factor” means the reciprocal of the probability of selection). An important distinction is made between forecasts and estimates. A forecast is a projection of a parameter or value that does not yet exist. For example, CSA forecasts crop yields and production early in the growing season and well before harvest. In contrast, an estimate is a measure of something that exists, like planted hectares or livestock inventory. CSA estimates area cultivated, yield and production through crop cutting exercises at the harvesting time. Meher seasons crop area and production forecasts are generally released in November whereas estimates are released in April.

CSA enumerators measure the area of nearly 600,000 crop fields every Meher season and on one thirds of them, crop cuttings to estimate yield is conducted

4. MoARD Data Generation System. The Ministry of Agriculture and Rural Development (MoARD) is the main Government institution responsible for policy and planning regarding agricultural development. This includes crops, livestock and all other aspects of agriculture including fisheries, beekeeping etc. MoARD has several subject matter departments such as Planning and Programming, Agricultural Extension Service, Crop Production and Protection Technology and Regulatory, Animal and Fisheries Development. All the subject matter departments are involved in some form of farm data collection and its utilization in their respective areas. For this purpose, a hierarchy of field level agencies is in place in every Region. MoARD data are collected subjectively at the lowest administrative level. Data are then aggregated at these low administrative levels and transmitted upwards from Peasant Associations (PA) to Woredas, Zones and Regional levels. Crop area and crop production data are generated in this way. For administrative purposes, the country of Ethiopia is divided into 11 regions (9 rural and 2 urban), with considerable regional heterogeneity. Regions are divided into zones and zones are divided into woredas. There are over 700 woredas in the country. Within a woreda, there are a number of Peasant Associations (PA's) in rural areas, and Urban Dweller associations in urban areas. These associations form the lowest administrative units in the country. In every PA, there are Development Agents (DA's) who are at the lowest level of Government agency. They are the main links between farmers and all development programmes of the Government. In fact, three DA's (one for crops, one for livestock and the other for natural resources) are supposed to be in place in every PA. Wherever there are less than three DA's, the work is shared among the existing ones. There are supervisors located at the woreda level, who also assist and support the DA's, wherever needed. The approach in data collection at MoARD is not standardized. Regions have a substantial amount of autonomy. Therefore, some of the practices regarding data collection and maintenance vary from region to region and often within a region as well. Even the data collection format differs. In every PA, some contact farmers (normally one contact farmer for a group of 20 to 30 households) are identified. These contact farmers are nodal points and are the link between DAs and the individual farmers. They are in touch with farmers on a routine basis and are supposed to have precise ideas about land cultivated under different crops by individual farmers. Information provided by the contact farmers is the starting point for the crop area statistics in the "subjective" approach of MoARD for generating crop production forecasts at the woreda level. The contact farmers report the crop area to the respective DA's. Information on area sown and crop prospects are reported fortnightly and this becomes an important factor of the crop forecast. The DA consolidates these figures for the entire PA and transmits further to woreda level. Information flows from farmers to farmers focal points to DAs to woreda, zone (where zone offices are active) to regional levels are often dislocated. In order to address information gaps, woreda, zonal and the regional BoARD are fielding crop assessment missions for each cropping season.

These missions are visiting fields, talking to DAs and farmers in order to generate crop area and production figures. As a result, it appears that DAs observations are not systematically utilized.

5.Complementarities and Differences.Considerations on aggregation and reporting levels are of importance in Ethiopia, especially when comparing CSA probabilistic sample survey and MoARD subjective method. The below conceptualize the differences in both approaches. CSA implements a statistically sound methodology to generate mandated estimates at Federal, Regional and Zonal levels whereas MoARD generate estimates at the critical Woreda level. To illustrate the situation, let us consider a hypothetical situation with nine small administrative areas (e.g. woreda) sum to a zone. Here, our objective would be to estimate wheat production. The diagram below shows the model.



There are two approaches of estimating wheat production in this model:

1.One could visit and conduct subjective observations and “guesstimate” wheat production in each of the nine woredas, or: 2.One could set up a sampling frame, select a sample of three units (farmers or small geographical area) at random and collect data to generate an estimate for each of the nine woreda. Which of the two approaches will be more accurate? It depends on the reporting level (woreda or zone). Certainly if the sample is three (or less) at the woreda level, the first system described above can be more accurate within each of the nine areas. A person’s judgment can be closer to a target value than a probability sample of three when there is high variability between the units. However the estimate at the zone level is a totally different story. The second method will indeed be more accurate because we have a larger sample of 27 units (3 lots of 9) where data has been collected. With a sample of five and above, probability sample approaches are closer to a target value than a subjective approach. A probability sample design shall always begin at the highest level of aggregation as those estimates are based on the largest number of samples. The published estimates should always begin at the national level and the most accurate estimates are broken into regional, district and even woredas level estimates.

Non-probabilistic subjective estimates starts are the lowest administrative level. But, if these estimates are aggregated at national level, these are usually far from accurate. The advantage of statistics as estimates is that sampling errors compensate when they are added so the more sample there are, the more accurate the estimate becomes. This is in striking contrast to non-probabilistic subjective estimates in which errors accumulate, and estimates become less accurate when estimates are added together. Therefore, although these subjective estimates have values at the smaller administrative levels, these shall not be aggregated.

6.Generating Crop Production Estimates at Woreda Level.Woreda level estimates are important to the Government of Ethiopia, particularly for food security and early warning purposes. One way to provide these woreda level estimates is to increase the sample size. The sample would have to increase to at least 30 in each woreda and there are far too many woredas (700+) for this to be viable. The total sample size would be above 20 000 EAs which is the approximate number of EAs sampled for the 2001 National Agriculture Sample Census Enumeration (NASCE). Cost, time and accuracy would be affected adversely if the sample size were to be increased to accommodate woreda level estimates.


Generating woreda level data every cropping season is a required by the Government of Ethiopia. MoARD estimates have value at the critical woreda level but these shall not be agregated at higher administrative level

7.Small Area Estimators - SAE. A cost-effective solution is to use auxiliary information from other sources and apply the technique of small area estimation (SAE). Fortunately, at present there are a number of reliable techniques available for SAE and there are sources of auxiliary information. These include MoARD area and production estimates at woreda level and the 2001 National Agriculture Sample Census Enumeration (NASCE).The terms small domain or area typically refer to the part of a population for which reliable statistics of interest cannot be produced due to small sample sizes in the domain. These domains are not always geographical but in the case of Ethiopia and this discussion, they are geographic. These may as well be small parts of the population based on cross classifications of people such as sex, age, qualification etc. The most widely used statistical method developed in early stages is Synthetic Method of Estimation. In our case, estimates generated at zone level are disaggregated to produce woreda level estimates, using Ministry of Agriculture estimates as auxiliary variable. The first step in applying SAE is to generate direct estimates for selected crops at the woreda level from the annual CSA agriculture survey. These estimates are used as dependent variables. Then, corresponding auxiliary data from Regional Bureau of Agriculture and Rural Development (BoARD) and the estimates from National Agriculture Sample Census Enumeration (NASCE) (2001) are used as auxiliary independent variables for regression fitting. This is carried out in order to choose suitable variables for model fitting. In this case, the R2 value was obtained as 0.6 and regression could be fitted with the corresponding auxiliary variables (BoARD and NASCE datasets). The SAE developed for Ethiopia combined the Empirical Best Linear Unbiased Predictor (EBLUP) and Ratio Synthetic Estimators models for better outputs. Procedures for developing SAE estimates are written and appropriate software fully developed and automated by CSA professionals. The SAE technique chartered the way for a collaborative work between CSA and MoARD. Using the SAE technique, both institutions are collaborating to produce reliable estimates from federal to woreda levels.

Using SAE technique, CSA and MoARD are collaborating in producing reliable estimates from Federal to Woreda levels.

8.Standardizing Data Collection Format.Until recently, MoARD data generation system did not have documented concepts and definitions and reporting schedules varied from Region to Region and Zone to Zone. Data items of agriculture have to be distinctly defined and identified. The correct way of stating data items and related terms is a prerequisite for making standards and definitions for the collection and compilation of agricultural data. The purpose of using standard concepts and definitions is not only to provide quality data but also to ensure that the right items are enumerated and measured accurately to reflect the agricultural situation. With the support of the project, the CSA and MoARD agreed upon standard concepts and definitions for the purpose of MoARD/BoARD agriculture data generation. This also includes a standard reporting format from DAs to woreda Bureaus of Agriculture and Rural Development (BoARDs). In September 2008, a comprehensive training on these concepts was given to each woreda BoARD by CSA. Two senior professionals from each woreda underwent “Training of Trainers” and then went on to provide ‘on-the-job” training of DAs on how to complete the standard forms. The training was coordinated by the project Crop Acreage Task Force. These standardization efforts aimed at increasing the R2 value of auxiliary data generated by MoARD/BoARD. This will allow for higher accuracy Woreda level estimates by applying SAE technique.

Standardisation of MoARD data collection format will allow for higher accuracy Woreda level estimates by applying the Small Area Estimation technique.

9.Looking beyond the Project Timeframe.By harmonizing agricultural statistics, the project is strengthening the fundamentals on which food security information systems and analysis are based. The project will have a long-lasting impact on the way food security related data is collected, managed, disseminated and analyzed. However, a three years project is too short to fully develop all required technologies narrated above and particularly to consolidate policy and organisational implications these technologies are having on project partner institutions. The Government of Ethiopia requires follow up support to build on the project achievements. Consolidating Consensus on harmonized Agriculture Statistics.With support from the project, a suitable methodology for estimating crop production was successfully defined as was harmonization of an agriculture statistic system at the federal level. However, implementing the harmonized methodology nation-wide requires a concerted effort between CSA and MoARD/BoARD at regional and field levels. The capacity building exercise for MoARD/BoARD is not a one shot exercise and training programmes should continue. In addition, the SAE technique should be further refined through other reliable auxiliary variables and this procedure sustained in order to produce woreda level estimate every year. Furthermore, advocacy to institutionalize the use of these datasets by institutions conducting food security analysis is highly recommended.

For additional information, please contact:
Samia Zekaria, CSA Director General: samiaz@ethionet.et
Wondirad Mandefro, MoARD Director Extension: wondirad_mandefro@yahoo.com
Raphy Favre, FAO Project Chief Technical Adviser: raphyfavre@gmx.net

photos by Raphy Favre design by Donu Iordache printed in Ethiopia