Master Sampling Frames for Agriculture

Supplement on selected country experiences

Brazil, Bulgaria, China, Ethiopia, Georgia, Guatemala, Nepal, Rwanda, USA and USDA experiences in Nicaragua, Nigeria and Tanzania

January 2018
Master Sampling Frames for Agriculture

Supplement on selected country experiences

Brazil, Bulgaria, China, Ethiopia, Georgia, Guatemala, Nepal, Rwanda, USA and USDA experiences in Nicaragua, Nigeria and Tanzania

# Table of Contents

Preface............................................................................................................. 4  
Acknowledgments............................................................................................. 6  
Overview and general outlines of country papers – Naman Keita........... 7

## COUNTRY PAPERS

- **Brazil** – Flavio Bolliger............................................................................. 10  
- **Bulgaria** – Mariana Toteva and Diana Atanasova................................. 29  
- **China** – Xinhua Yu.................................................................................. 44  
- **Ethiopia** – Aberash Tariku Abaye.......................................................... 62  
- **Georgia** – Maia Guntsadze and Vasil Tsakadze................................. 77  
- **Guatemala** – Marino Barrientos............................................................. 93  
- **Nepal** – Ambika Bashyal......................................................................... 111  
- **Rwanda** – Sebastien Manzi..................................................................... 115  
- **USA** – Sarah Hoffman............................................................................ 171  
- **USDA experiences in Nicaragua, Nigeria and Tanzania**  
  Sarah Hoffman.............................................................................................. 177
Preface

This supplement on selected country experiences to the Handbook on Master Sampling Frames for Agriculture has been prepared within the framework of the Global Strategy to Improve Agricultural and Rural Statistics (Global Strategy). The Global Strategy is an initiative endorsed in 2010 by the United Nations Statistical Commission. It provides a framework and a blueprint to meet current and emerging data requirements and the needs of policymakers and other data users. Its goal is to contribute to greater food security, reduced food price volatility, higher incomes and greater well-being for rural populations, through evidence-based policies. The Global Strategy is centred upon three pillars: 1) establishing a minimum set of core data, 2) integrating agriculture into national statistical systems, and 3) fostering the sustainability of the statistical system through governance and statistical capacity building.

As indicated in the Global Strategy’s foundational document, the implementation of the second pillar (integration of agriculture into the national statistical system): “begins with the development of a master sampling frame for agriculture that will be the foundation for all data collection based on sample surveys or censuses.”

The Action Plan to Implement the Global Strategy prioritized the preparation of the Handbook on Master Sampling Frames (MSF) for Agriculture Frame Development, Sample Design and Estimation to provide statisticians in countries with practical guidelines.

The Handbook on MSF was prepared by a core team of five senior consultants and experts, and published in December 2015. The process of developing the Handbook on MSF included a dedicated expert meeting organized by the Global Office in Rome in November 2014 with international, regional and national experts. Experts from several countries prepared detailed technical papers that illustrate their diverse experiences in building and using sampling frames for agriculture surveys, and it was recommended that these papers be consolidated into a practical supplement to the Handbook on MSF, which includes in its annex a short summary of the country papers.

This supplement to the Handbook on MSF is intended to provide practical examples on how sampling frames can be developed and used as master sampling frames in different country contexts given the diversity of country situations and resources.
The country experiences included in the supplement cover a wide range of countries worldwide. The papers were prepared by senior national experts from Brazil, Bulgaria, China, Ethiopia, Georgia, Guatemala, Nepal, Rwanda and the United States of America, following a standard outline provided by the Global Office of the Global Strategy.

The supplement to the Handbook focuses on the practical aspects of building a sampling frame and using it as an MSF, and is intended to be a companion document to the handbook, which provides more details on the methods. The supplement reviews countries’ experiences with regard to: 1) details of methodology used for building a sampling frame and survey design, including type of frame used and why, the process of developing the frame, sampling design, formulas for estimators and variances; 2) results obtained and precision of estimates; 3) resources mobilized and other costs; 4) how the data collection was organized; and 5) issues and challenges and lessons learned. The purpose is to allow the reader to make the comparison and facilitate the application of the most appropriate method for frame building and using.

The supplement is conceived as a “living document and could be revised when more relevant country experiences are available. It will be made available at http://www.gsars.org/category/publications/.
Acknowledgments

The experiences described in this Supplement to the Handbook on Master Sampling Frame for Agriculture was prepared by a group of nine senior agricultural statisticians from selected countries. The country papers were edited and consolidated by Naman Keita, Senior Statistician Consultant of the Global Strategy’s Global Office. But care has been taken to preserve the integrity and individuality of each report as initially prepared by the author.

The country papers were prepared by the following authors:

- **Brazil**: Flavio Bolliger, Instituto Brazilian do Geografia e Estatistica.
- **Bulgaria**: Mariana Toteva, and Diana Atanasova, Agrostatistics Department of the Ministry of Agriculture and Food.
- **China**: Yu Xinhua, Division of Informatics and Methodology Department of Rural Surveys National Bureau of Statistics.
- **Ethiopia**: Aberash Tariku Abaye, Central Statistics Agency.
- **Georgia**: Maia Guntsadze and Vasil Tsakadze, National Statistics Office of Georgia.
- **Guatemala**: Marino Barrientos, Ministerio de Agricultura, Ganadería y Alimentación.
- **Nepal**: Ambika Bashyal, Central Bureau of Statistics.
- **Rwanda**: Sebastien Manzi, National Institute of Statistics.

Valuable inputs and comments were also provided during and after a dedicated high-level expert meeting on Master Sampling Frames organized by the Global Office at FAO headquarters in November 2014.

This publication was prepared with the support of the Trust Fund of the Global Strategy, funded by the United Kingdom’s Department for International Development, and the Bill and Melinda Gates Foundation. The World Bank and the Joint Research Center of the European Union also provided financial and technical support towards preparing the publication.
Overview and General Outline of Country Papers

by Naman Keita

The Global Strategy’s Handbook on Master Sampling Frames (MSF) specifies that “countries follow a variety of practices when building sampling frames for agriculture and in using these frames as MSFs”. These practices are based on the three types of frames discussed in the handbook: area frames, list frames and multiple frames, which combine list frames and area frames.

This supplement to the Handbook on MSF provides a review of ten selected country examples on how these concepts are applied in practice in a specific country context to build and use a sampling frame for agricultural censuses and surveys.

The country papers that follow show that many countries have used list frames and area frames successively or jointly in a multiple frame. The change in the type of frame used is often due to the rapidly changing agricultural structure and statistical environment (e.g. in China, Rwanda, Bulgaria) and the need to adapt to the new situation.

Several countries use population and housing census (PHC) data and/or agricultural census data for building their sampling frame for subsequent or current surveys. Census enumeration areas (EAs) are generally used as primary sampling units (PSUs) or grouped into PSUs to constitute the main basis for MSFs based on list frames (e.g. Ethiopia, Nepal, Georgia, Bulgaria).

PHCs – with an agricultural module or conducted jointly with agricultural census – provide a more suitable source for building a list frame-based master sampling frame because relevant auxiliary variables can be obtained for an efficient frame building and sample design. When a PHC is followed by a complete enumeration agricultural census, the PHC can allow the identification of agricultural households, which can provide the basic frame for the census of agriculture. Then, during the census of agriculture, relevant auxiliary variables – such as quantitative values for cultivated land and/or area of crop and livestock inventories – can be obtained. This additional information can be used for
building an effective frame and efficient sample design for agricultural production surveys.

A major issue mentioned in all country experiences using this type of list frame is its rapid obsolescence and the need to develop some way of maintaining and updating the list between agricultural census cycles. List frames also seem prone to undercoverage and all efforts should be made to minimize this risk.

Regarding the use of area frames, the two approaches mainly used are sampling segments (square segments or segments with physical boundaries) and points. The country papers show that the construction of an area frame involves the important use of cartographic materials and satellite images, and image interpretation capacity is needed. Therefore, the teams developing the area frames should preferably include both survey statisticians and specialists in geosciences, image processing and geographic information systems. Building an area frame requires the mobilization of substantial initial human and financial resources, but the cost of conducting surveys and maintaining the frame decreases considerably (e.g. Bulgaria).

Some of the main advantages highlighted in the papers for area frame-based surveys include: 1) complete coverage of the target population; 2) reliability of statistical data, particularly area-related data; and 3) the possibility of collecting data through direct observation, thereby eliminating the subjectivity of farmers’ declarations. When same points or segments are observed every year, area frame surveys can be a powerful tool that not only estimates the area and sequence of crops, but also provides information on changes in land cover and land use as a result of structural changes.

Area sampling frames are, however, considered to have a major weakness for items that are not proportionately associated with cultivated land use, such as specialty or rare crops. Similarly, if there are relatively large operations that have a large impact on the item of interest, an area frame alone may not be suitable. The method is also considered to be less representative for small areas and for crops that are usually grown on small farms, such as tobacco and vegetables, and for orchards and vineyards. It can also generate an error of classification of the land use for categories whose use is difficult to determine by direct observation, such as permanent grassland, recently abandoned agricultural area, and fallow land. Other difficulties mentioned with using an area frame include:

- difficulty in localizing the selected segments;
- deficiencies in roads and paths, making it difficult to access selected segments in several cases;
difficulty in localizing respondents;
difficulty in determining the existence of the household or headquarters; and
other practical issues, such as farmers living far away from holdings, deciding whether to include the farm in a segment or not in case of a closed segment.

In most countries a multiple frame approach is adopted to address some of the deficiencies of a list frame and area frame. In many countries, a list of large and specialty operations — or for some items and variables such as livestock and rare crops — is combined with an area frame covering small holdings. This is particularly true in a country with mostly small household types of agriculture but also some large agricultural operations.

Finally, as specified in the Handbook on Master Sampling Frames, country practices indicate that a careful analysis of the country situation is needed — such as resources, materials available, institutional support, the scope of the statistical system and survey objectives — to ensure that the options selected are the most suitable for that country.

One of the lessons that appear from the country experiences is that the managers of the countries’ agricultural statistics systems need to constantly review and adapt the methods used, including frame building and using, to the rapid evolution and transformation of agricultural sector and to the improvements in technology, tools, methods and administrative system, particularly in developing and transition countries.

The subsequent sections present details of country experiences as described by the national experts.
For 14 years, the Brazilian Institute for Geography and Statistics (IBGE) has been conducting an agricultural survey based on probabilistic sampling, referred to as the Harvest Forecasting and Monitoring Survey (PREVS, in Portuguese). The first edition of the survey referred to the 1986/1987 harvest while the latest edition referred to the 1999/2000 harvest. This paper describes aspects of the methodology used in the PREVS, emphasizing the frame system used, estimators and level of accuracy obtained, main difficulties and lessons learned. Some reasons for the discontinuation of the survey are also discussed. In addition, basic aspects of IBGE’s proposals, at the time of preparing this paper, to re-start the implementation of continuous agricultural surveys by probabilistic sampling are presented, emphasizing solutions regarding the agricultural sample frame.

**Introduction**

The Harvest Forecasting and Monitoring Survey (known as PREVS in Portuguese) was instituted by the Brazilian Institute for Geography and Statistics (IBGE) in the mid-1980s. At its implementation, and up to 1993, the survey was financially supported by Banco Internacional para Reconstrução e Desenvolvimento (BIRD). PREVS followed approximately the methodology that was used by the National Agricultural Statistics Service of the United States Department of Agriculture in its “June Enumeration Survey and December Enumeration Survey”.

The PREVS publication\(^1\) specifies that “the main purpose of the survey is to provide statistic information regarding harvests, necessary for the evaluation and agricultural planning of the Federation Units which are included in the project, through the objective data collection methods and the use of probabilistic sampling, which allows the association of a confidence interval to the final results” (free translation from IBGE, 1998).

Although the survey produced various information of agricultural interest on economic data, land use, use of fertilizers and pesticides, livestock data and other data, it is important to note that the project’s main goal was “to collect information regarding the harvest forecast ...” (IBGE, 2005:27).

**Methodological characteristics of PREVS**

The survey followed an area sample frame design: stratification according to land use and a systematic selection in a single stage, with equal probability and with no replacement.

**Building and using a sampling frame**

The area sample frame consisted of strata of land use established according to the rate of cultivated land, or by the predominance of crops, divided into counting units (CUs) that were subdivided into area segments, the survey’s sampling unit.

The land-use strata were constructed through the application of satellite image interpretation techniques, making use of images from the TM/LANDSAT sensor (scales 1:100 000 or 1:250 000), jointly with the use of topographic maps, using as its base the information deriving from IBGE’s agricultural census, the Systematic Survey of Agricultural Production (LSPA, in Portuguese), Municipal Agricultural Survey (PAM, in Portuguese) and the Municipal Livestock Survey (PPM, in Portuguese). Taking into account the structural and spatial distribution of the agricultural activity in the surveyed states, the area strata were established as described in Annex 1 and represented in Annex 2.

With the purpose of increasing the model’s efficiency, a second stratification level was adopted, named geographic substratification, based on conglomerate analysis techniques, with the restriction of geographic continuity (the sub-strata defined should be a continuous surface). Substrata were constructed to contain the same number of segments, except for the last substratum of the stratus, in cases where it was not possible to achieve such equality for all of them. The substrata should still contain a number of segment multiples of five in order to facilitate the sample rotation by 20% a year, and to avoid respondent burden. Such sample rotation, however, was never put into practice.

The CUs were defined with permanent physical boundaries identified and with the purpose of avoiding the division of the entire federal unit into segments. The

---

2 The description of the methodology was compiled from IBGE 2005 and FAO 1998.
3 São Paulo, Paraná, Santa Catarina and Distrito Federal.
segment is the result of the last stage of the sample frame construction. The segment is limited by the terrain’s identifiable permanent physical boundaries and constitutes the survey’s sampling unit. It is important to point out that the CUs did not constitute the selection unit, being present merely as an artifact with the objective of operational simplification. To this purpose, an average size of the segment was pre-established, and determined by taking into consideration the characteristics of each stratum. The ratio between the area of each stratum and substratum and the average size of the segment, defined the quantity of segments contained in each one, and in each CU.

The size of the sample was calculated based on the variance estimates for some of the variables of interest obtained in a pilot test. The sample allocation was proportional to the farming area in each land-use stratum, with the application of a systematic selection of segments in each substratum.

**Estimators**

There are three different methods to associate the information units to the sample unit: the closed segment, if the information unit was the area of holding; an open segment and weighted segment, if the information unit was the establishment. Thus, three kinds of estimators were considered:

- Close segment estimator
- Open segment estimator
- Weighted segment estimator

The estimate for the total of the variable X was obtained as follows:

$$
\hat{X} = \sum_{h=1}^{L} \sum_{l=1}^{L_h} e_{hi} \sum_{j=1}^{n_{hij}} l_{hij}
$$
Where:

\( \hat{X} \) is the estimate for the total of the variable \( X \)

\( L = \text{number of land usage strata} \)

\( L_h = \text{number of substrata, of the } h \text{ stratus} \)

\( e_{hi} = N_h / n_{hi} \), expansion factor for each \( j \) segment, in the \( i \) substratum, in the \( h \) stratus.

\[ t_{hij} = \sum_{k=1}^{m_{hij}} t_{hijk} \]

value of the variable in the holding area \( k \), in the \( j \) segment, substratum \( i \), stratus \( h \)

\( m_{hij} = \text{number of holding areas in segment } j, \text{ in the substratum } i, \text{ stratus } h \)

With \( t_{hijk} = x_{hijk} \), for the closed segment estimates

\[ = x_{hijk} \cdot a_{hijk}, \text{ for the open segment estimates} \]

\[ = x_{hijk} \cdot w_{hijk}, \text{ for the weighted segment estimates} \]

The factors \( w_{hijk} \) and \( a_{hijk} \) are defined as:

\( w_{hijk} = \text{weight used in the area of holding } k, \text{ segment } j, \text{ and stratus } h \)

\( a_{hijk} = 1 \), if the establishment \( k \)'s headquarters or household belong to segment \( j \); and

\( a_{hijk} = 0 \), if the establishment \( k \)'s headquarters or household doesn’t belong to segment \( j \).
The formula for the variance is similar for the three kinds of estimators based on substrata and using the successive difference method:

\[ \text{var}(\hat{X}) = \sum_{h=1}^{H} \sum_{i=1}^{I^h} \frac{N^2_{hi}}{n_{hi}} \left[ (i - f_{hi})/n_{hi} \right] \cdot s^2_{hi} \]

Where:

\[ f_{hi} = \frac{n_{hi}}{N_{hi}} , \text{ sampling fraction for each } j \text{ segment, at substratus } i, \text{ in the } h \text{ stratus.} \]

\[ s^2_{hi} = \sum_{j=1}^{n_{hi}-1} \left( t_{hij} - t_{hi(j+1)} \right)^2 / 2(n_{hi} - 1) \]

Estimates on land use and pig farming were obtained by the closed segment method. In the cattle estimates, the weighted segment method was used, the weight factor being the ratio between the pasture area within the segment and the total area of the establishment. Regarding the amount of seeds, the weight factor was the quotient between the cultivated areas within the segment and the segment’s total area.

**List frame for special holdings and multiple frame estimators**

Based on information from the 1985 agricultural census, several lists of holdings that represent a large percentage of the total of the variable were constructed. These lists, including a relatively small number of holdings, are called Lists of Special Holdings and are updated every year.

For variables with a distribution — such that a large percentage of the total is concentrated in a relatively small number of holdings — multiple frame estimators were used in order to improve the accuracy of the direct area sample estimators.

For a given variable, the multiple frame estimator is the sum of the estimators from both samples, the area sampling estimator, and the list sample estimator based on the list frame of special holdings.

After the agricultural year 1994/1995, the multiple frame technique was no longer used because the list became obsolete and there was no updated list frame.
Results obtained and accuracy of estimates

PREVS provided: annual estimates of planted and harvested crop areas, and areas intended for harvest; potential and actual yields; and production for the main summer crops (maize, soyabean, bean, and others), production of cassava and coffee, and cattle, establishment characteristics. Some rounds of PREVS included other data such as the use of pesticides, labor force, income and expenditure. PREVS produced statistics for the State of Paraná for the harvest years of 1986/1987 to 1999/2000 and Special Survey of Coffee in 1997 and 1998; for Santa Catarina and the Federal District from 1987/1988 to 1994/1995; and for São Paulo from 1989/1990 to 1995/1996. Due to the operations of the population census and agricultural census, PREVS’s ordinary data collection was not performed in the years 1991/1992 and 1996/1997. However, the same statistical infrastructure and survey methods were employed during the Special Survey of Coffee in Paraná in 1997 and 1998.

Table 1. Distribution of the harvest estimates and other variables published by PREVS by level of accuracy, 1987–2000

<table>
<thead>
<tr>
<th>Indicator/Concept</th>
<th>Coefficient of variation (CV)</th>
<th>Crops estimates</th>
<th>Estimates for other topics¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>% acum.</td>
<td>%</td>
</tr>
<tr>
<td>Z Exact</td>
<td>Zero</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>A Great</td>
<td>Up to 5%</td>
<td>11.3</td>
<td>11.3</td>
</tr>
<tr>
<td>B Good</td>
<td>Above 5 up to 15%</td>
<td>35.1</td>
<td>46.4</td>
</tr>
<tr>
<td>C Reasonable</td>
<td>Above 15 up to 25%</td>
<td>19.8</td>
<td>66.2</td>
</tr>
<tr>
<td>D Little accurate</td>
<td>Above 25 up to 50%</td>
<td>21.4</td>
<td>87.7</td>
</tr>
<tr>
<td>E Inaccurate</td>
<td>Above 50%</td>
<td>12.3</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Source: IBGE, Harvest Forecasting and Monitoring Survey (PREVS)

Note: Estimates with indications of level of precision for Paraná, Santa Catarina, São Paulo e Distrito Federal, several years.

¹ Livestock, agricultural practices, labor force, economic results and others.

It appears that 66.2% of the crop estimates published by PREVS had shown an accuracy considered to be reasonable (C indicator – CV from 15% up to 25%), or better. Among the other variables surveyed, great or good accuracy indicators were found in 73.5% of cases (Table 1). In general, results of the average yield of farming, livestock numbers and land use, in addition to the planted area of some of the most common crops of each state, showed better precision: great (indicator A – CV less than 5%) and good (indicator B – CV between 5% and 15%).
Resources mobilized

In the headquarters, two technical teams were formed to address the project. One team consisted of 14 technicians from the Board of Geosciences who were responsible for processing images, and building and maintaining the area frame and geographic information system (GIS) for the survey programme. The other team consisted of 11 technicians from the Board of Surveys who were responsible for determining the statistical and sampling procedures, manual and computer data editing, imputating of missing or unreliable data, tabulating and analysing results, and disclosure.

In each state, a survey coordination unit was created, comprising one coordinator, the supervisors and the enumerators. IBGE maintains an extensive network of field offices in each of the surveyed states. These state statistical offices provide the infrastructure for the survey, logistics for sending and receiving survey materials, vehicles and communications.

The field data collection for each survey was undertaken by 20 supervisors and 200 enumerators.

The building of the area frame involved the use of the following cartographic materials:

- Statistical municipal maps (MMEs) prepared for census purposes. These show each municipality divided into enumeration areas. In order to improve the stratification process, statistical agents were asked to mark on each MME the areas for more crops. The scale used in the MMEs were usually 1:100 000.

- Political maps and land-use maps.

- Topographic maps. Generally, the scales used in topographic maps were 1:50 000 and 1:100 000. In the Distrito Federal, however, the maps used a scale of 1:25 000 and dated back to 1985.

- Satellite imagery TM/Landsat-V on paper, at scales of 1:100 000 and 1:250 000.

- Index map of TM/Landsat covering.

- Planimetric conjugates, produced on clear acetate, at the same scale as the TM/Landsat images. These were used for visualization and
geographical location of the interpreted patterns. The planimetric conjugate is the result of superimposing the planimetric layer on the topographic map. It originated from a map at scale 1:50 000, later reduced to 1:100 000.

- Aerial photos covering the selected segments. Aerial photos were available from companies in the original flight scale. Photo enlargements to 1:10 000 were used for the annual field data collection.

- Photo indexes. For example, the photo index of the Distrito Federal at scale 1:120 000, dated back to 1986. (FAO, 1998:12).

An Intergraph and Maxi-Cad workstation, an IBM 3090 computer, DB2 language and SAS software were also used.

**Difficulties and challenges**

Team members reported common difficulties in conducting fieldwork, including difficulty in localizing the selected segments; deficiencies in roads and paths, making it difficult to access selected segments in several cases; difficulty in locating respondents; difficulty in determining the existence of a household or headquarters or, alternatively, in verifying if the largest part of the establishment belonged to the segment (in cases where the producer lived in the city). It was also difficult to update the strata for the area frame and prepare fieldwork materials, and there was a lack of sources for the list frame’s update.

**Why was the survey discontinued?**

IBGE (2005) reported on the shortage of technical teams of the regional offices and on the difficulty in establishing partnerships with local entities to continue the survey (IBGE, 2005:28). It is evident that at some point there was a lack of human and financial resources and institutional support to sustain the survey. However, there are deeper reasons for such an outcome, and these pertain to the objectives, implementation strategy, and the relevance of the results provided by the survey.

An important deficiency that was pointed out was the fact that the survey did not have nation-wide coverage. It started as a pilot in Paraná State and was expanded throughout the years to other Brazilian states, including São Paulo, Santa Catarina and Distrito Federal, but no others. Such coverage was considered insufficient, given the IBGE board’s position to carry out nation-wide surveys.
Another important limitation was the timeliness of the survey. The survey focused on harvest forecasting and monitoring, which resulted in only one round of data collection a year, because the survey centered only on the summer harvest. Data collection initially took place in the first months of the year (January to March), and the results were released in the middle of the year (between June and July) when the summer harvest was almost entirely harvested and sold. In the agricultural year of 1993/1994, data collection was brought forward to the months of October and November, which was the planting period of the main crops surveyed. This expediency also showed many limitations due to the large “planting window” that is observed in Brazilian agriculture. A part of the information collected referred to the planting intentions and, the information collected on yield corresponded to expectations and not to observed results.\textsuperscript{4}

Paradoxically, it is believed that an additional limitation comes from the survey’s own quality, which was based on the best and most sound methods, which were in turn, derived precisely from utilizing probabilistic sampling and producing information on the accuracy of estimates. Indeed, as illustrated in charts 1 and 2, the survey served mostly to confirm other estimates obtained by subjective methods when shown inside of the confidence interval of the PREVS’s results.

Chart 1 compares the estimates of PREVS and the Systematic Survey of Agricultural Production (LSPA) to the final official data published by the Municipal Agricultural Survey (PAM), for PREVS’s most accurate area data — soybean data in Paraná State — showing its time of publication. Throughout its survey period, PREVS indicated estimates consistently superior to those of LSPA. Besides the size and spread of the sample, which constituted the greatest effort of data collection, estimates were associated with a confidence interval of a medium amplitude, equivalent to 22% of its value. Chart 2 shows these relations for other similar cases. The central value of PREVS’s estimates for soy or any other crop was never adopted as official data, whether by the Coordinating Groups of Agricultural Statistics (GCEA in Portuguese) of each state, or by the Special Commission of Planning, Control and Evaluation of Agricultural Statistics (CEPAGRO in Portuguese) at the national level.

In addition to these considerations, there is the country’s long tradition of agricultural statistics production based on subjective surveys. The LSPA, with much less effort and cost, has been providing — for over 40 years — monthly estimates of planted area, harvested area, average yield and production for more

\textsuperscript{4} An Objective Yield Survey was proposed within the project’s scope, but was never implemented.
than 30 products and for all of the Federation Units, even if the generation of such quantitative information through subjective surveys was considered questionable\(^5\) in terms of its methodology. Generally, PREVS’s results were released long after LSPA’s estimates were already consolidated and agreed on between users and producers and used as official data.

**Lessons learned and new challenges**

A system of agricultural sample surveys continues to be an important aspiration to the improvement of Brazilian agricultural statistics. The National System of Agricultural Holding Sampling Surveys Project (SNPA) was under development within IBGE’s Coordination of Agriculture at the time of preparing this notes. Considering the experience with PREVS, it is envisaged to implement SNPA to cover the entire national territory although crop forecasting is not an objective. Its main survey, the Annual National Agriculture Survey (PNAG), was planned as a multipurpose and continuous survey, covering various aspects of social, economic and environmental issues. The proposed system includes a current survey — the Quarterly Production Survey, or PNPA — that collects quarterly data on agricultural production but does not aim at forecasting and monitoring information.

The design of a master sample frame for SNPA, including a list frame and area frame, was planned through the traditional approach, using enumeration areas (EAs) and information from agricultural censuses. Instead of using segments such as selection units, the model has as its primary sampling unit the EA (or a group of contiguous EAs) and selects, in a second stage, the agricultural establishment.\(^6\)

---

\(^6\) More detailed information on the master frame project and sample design can be found at Bolliger et al. 2012 and Freitas et al. 2013.
Chart 1: Soya bean harvest area estimates (ha) - Paraná, 1987-1995

Sources: IBGE, Agricultural Production Systematica Survey (LSPA); Harvest Forecasting and Monitoring Survey (PREVS), Agricultural Production Municipal Survey (PAM)

Chart 2: Production estimates, selected crops - Paraná, 1994-1995

Sources: IBGE, Agricultural Production Systematica Survey (LSPA); Harvest Forecasting and Monitoring Survey (PREVS), Agricultural Production Municipal Survey (PAM)
A question that remains regards the deterioration of the master sample frame over time, and the mechanism for updating it, given the fact that Brazil is a country of significant dynamism regarding land use. It is evident that a sample based on a master frame built from the 1996 agricultural census would underestimate agricultural activities in the states of Mato Grosso, Maranhão, Piauí and west of Bahia after ten years. And, the master frame would not properly reflect the movements of livestock towards the north of the country and even the substitution of pastures and other crops by sugarcane, in a state of secular agricultural occupation such as São Paulo.

The discrepancy of the last (2006) census data was one of the main reasons for IBGE’s decision in 2013 to postpone the implementation of SNPA, in order for it to happen alongside the implementation of the next (11th) Brazilian Agricultural Census that was planned for 2016 but later postponed to 2017.

Under these circumstances, the Coordination of Agriculture initiated studies on methodological alternatives to SNPA’s implementation, in particular, regarding the master sample frame for agricultural surveys.

One of the alternatives considered is quite radical and relies on advancements observed in the country’s administrative records system. In this alternative, the target population of SNPA would be redefined, becoming the set of units of the agricultural establishments under the domain of the agricultural producers formally registered with the federal or state public power.

In this case, the target population is more restricted than that of the agriculture census. This will exclude agricultural establishments whose primary function is leisure, housing or livelihood, or producers that have agriculture as a secondary activity for which the holders do not have to liaise with the government for funding or technical assistance, nor engage in the marketing of products to the point of seeking to observe the provisions of movement of goods. The relevance of these constraints is minimized by the current scope of the administrative records, in particular, the Pronaf register, which is directed at family farms and now has almost 5 million active records.

SNPA’s sample frame would consist of a list of rural producers. The Agricultural Sample Frame formed by combining the main sample frames maintained by the federal and state governments — which register agricultural producers — would be updated periodically and would form a list frame focused on SNPA’s general

---

7 National Program to Support Family Farming
surveys. Such an approach would present the following advantages, disadvantages and risks.

Advantages

- The reference system’s annual updating
- Independence regarding the census maps
- Independence regarding the agricultural census
- Low implementation cost.

Disadvantages

- More restricted target population (the same types of holdings are not covered by the administrative records).
- Does not meet the requirements for integration with the agricultural census (integration with census and survey is a desirable goal for the agricultural statistical system aiming at consistency of estimates and results).
- Greater dependence on the quality of external registers.

Risks

- Difficulty accessing external registers
- Alteration and discontinuity of external registers (the administration could change the content of the register).

Another alternative considered is using the results of the new Project of Land Use and Coverage, developed by IBGE’s Geosciences Directorate. The project’s main goal is to monitor changes in the use and coverage of land for the entire national territory, in regular periods (every two years), from the acquisition and processing of MODIS (moderate-resolution imaging spectroradiometer) images.
The project maps the following categories and classes of land use and coverage:

1. Anthropogenic non-agricultural areas
2. Anthropogenic agricultural areas
   a. agricultural area
   b. planted pasture
   c. agricultural area with mosaic
   d. forestry
3. Forest vegetation areas
   a. forest vegetation
   b. mosaics of forest vegetation and agricultural activity
4. Rural vegetation areas
5. Water
6. Sparsely or not vegetated open spaces

Such information is provided through a territorial grid for statistical purposes. For its construction, the Albres Equivalente Conical Projection was used in order to guarantee equally sized squares. The official set of grids has the following levels:
<table>
<thead>
<tr>
<th>Title</th>
<th>Dimension (km)</th>
<th>Code</th>
<th>Area (km$^2$)</th>
<th>Number of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRE500KM</td>
<td>500</td>
<td>2556</td>
<td>250,000</td>
<td>56</td>
</tr>
<tr>
<td>GRE250KM</td>
<td>250</td>
<td>2456</td>
<td>62,500</td>
<td>181</td>
</tr>
<tr>
<td>GRE100KM</td>
<td>100</td>
<td>2555</td>
<td>10,000</td>
<td>960</td>
</tr>
<tr>
<td>GRE50KM</td>
<td>50</td>
<td>2455</td>
<td>2,500</td>
<td>3076</td>
</tr>
<tr>
<td>GRE10KM</td>
<td>25</td>
<td>2355</td>
<td>625</td>
<td>14,395</td>
</tr>
<tr>
<td>GRE5KM</td>
<td>10</td>
<td>2454</td>
<td>100</td>
<td>88,395</td>
</tr>
<tr>
<td>GRE1KM</td>
<td>5</td>
<td>2354</td>
<td>25</td>
<td>351,802</td>
</tr>
<tr>
<td>GRE500M*</td>
<td>1</td>
<td>2353</td>
<td>1</td>
<td>8,861,508</td>
</tr>
<tr>
<td>GRE250M*</td>
<td>0.5</td>
<td>2253</td>
<td>0.25</td>
<td>35,446,032</td>
</tr>
<tr>
<td>GRE100M*</td>
<td>0.25</td>
<td>2153</td>
<td>0.0625</td>
<td>141,784,128</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>2252</td>
<td>0.01</td>
<td>886,150,800</td>
</tr>
</tbody>
</table>

*not implemented yet

Figure 2 shows the incorporation of land-use and coverage data in the territorial grid for statistical purposes, in which it is possible to recover area data on the categories of land use and coverage for each square.

Figure 2. Land use and coverage incorporated into the territorial grid for statistical purposes.
The utilization of an area frame based on this information — stratifying and selecting squares — requires the use of electronic devices and a global positioning system (GPS) for field limits identification, and includes the following advantages and disadvantages.

Advantages

- Defines segments of the same size
- Permits the selection of segments that are multiples of 1 x 1 Km
- Biannual update
- Does not rely on census data.

Disadvantages

- No visually identifiable boundaries
- Larger variability of the number of establishments in each segment
- Based on less accurate information than provided by the agricultural census.

The use of both alternatives requires surveys, studies and field experiments to evaluate viability and efficiency (which are still ongoing at the time of preparing this paper) with a focus on building the most appropriate agricultural master sample frame.
References


Annex 1

(IBGE, 2005:190–194)

Table 1.1
Brazil: area sampling frame design information
State of Paraná - 1991

<table>
<thead>
<tr>
<th>Strata definition</th>
<th>Survey area (km²)</th>
<th>Target segment size (ha)</th>
<th>Number of CUs</th>
<th>Total segments (Np)</th>
<th>Sample segments (Nq)</th>
<th>Segments in subunits</th>
<th>Number of subunits</th>
<th>Size of subunits</th>
<th>1/N</th>
<th>Counting unit size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrigated agriculture, 50% and more of cultivated area</td>
<td>29,814</td>
<td>1</td>
<td>1,754</td>
<td>29,808</td>
<td>155</td>
<td>5</td>
<td>29</td>
<td>4,166</td>
<td>220.8</td>
<td>110</td>
</tr>
<tr>
<td>Irrigated agriculture, 30% to 49% of cultivated area</td>
<td>19,062</td>
<td>1</td>
<td>1,130</td>
<td>19,128</td>
<td>85</td>
<td>5</td>
<td>17</td>
<td>1,125</td>
<td>223.9</td>
<td>110</td>
</tr>
<tr>
<td>Special maize - potatoes</td>
<td>2,194</td>
<td>2</td>
<td>81</td>
<td>1,089</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>547</td>
<td>36.7</td>
<td>20</td>
</tr>
<tr>
<td>Irrigated agriculture - between 15% and 49% of cultivated area, with pedunculiform</td>
<td>23,913</td>
<td>2</td>
<td>881</td>
<td>11,916</td>
<td>55</td>
<td>5</td>
<td>11</td>
<td>4,080</td>
<td>214.1</td>
<td>60</td>
</tr>
<tr>
<td>Irrigated agriculture - between 15% and 49% of cultivated area, without pedunculiform</td>
<td>54,307</td>
<td>2</td>
<td>1,900</td>
<td>23,181</td>
<td>165</td>
<td>5</td>
<td>33</td>
<td>902</td>
<td>158.0</td>
<td>40</td>
</tr>
<tr>
<td>Irrigated agriculture - less than 15% of cultivated area</td>
<td>29,862</td>
<td>4</td>
<td>540</td>
<td>7,115</td>
<td>35</td>
<td>5</td>
<td>7</td>
<td>2,875</td>
<td>214.6</td>
<td>110</td>
</tr>
<tr>
<td>Non-irrigated land</td>
<td>29,312</td>
<td>8</td>
<td>735</td>
<td>6,957</td>
<td>20</td>
<td>5</td>
<td>4</td>
<td>1,334</td>
<td>249.5</td>
<td>170</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>104,797</strong></td>
<td><strong>-</strong></td>
<td><strong>6,916</strong></td>
<td><strong>90,540</strong></td>
<td><strong>520</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.4
Brazil: area sampling frame design information
State of São Paulo - 1991

<table>
<thead>
<tr>
<th>Strata definition</th>
<th>Survey area (km²)</th>
<th>Target segment size (ha)</th>
<th>Number of CUs</th>
<th>Total segments (Np)</th>
<th>Sample segments (Nq)</th>
<th>Segments in subunits</th>
<th>Number of subunits</th>
<th>Size of subunits</th>
<th>1/N</th>
<th>Counting unit size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrigated agriculture, 80% and more of cultivated area</td>
<td>1,116</td>
<td>0.5</td>
<td>816</td>
<td>10,237</td>
<td>36</td>
<td>3</td>
<td>12</td>
<td>853</td>
<td>384.56</td>
<td>17</td>
</tr>
<tr>
<td>Irrigated agriculture, 60% and more of cultivated area</td>
<td>17,015</td>
<td>1</td>
<td>1,640</td>
<td>17,027</td>
<td>90</td>
<td>30</td>
<td>11</td>
<td>811</td>
<td>243.34</td>
<td>18</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>21,813</td>
<td>2</td>
<td>897</td>
<td>11,292</td>
<td>15</td>
<td>5</td>
<td>7</td>
<td>1,428</td>
<td>323.51</td>
<td>18</td>
</tr>
<tr>
<td>Cattle</td>
<td>2,865</td>
<td>2</td>
<td>143</td>
<td>1,429</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1,429</td>
<td>287.00</td>
<td>17</td>
</tr>
<tr>
<td>Between 50% and 59% of cultivated area, plus pasture and woodland pasture and large fields</td>
<td>13,007</td>
<td>2</td>
<td>670</td>
<td>7,996</td>
<td>65</td>
<td>5</td>
<td>13</td>
<td>467</td>
<td>112.68</td>
<td>41</td>
</tr>
<tr>
<td>Between 30% and 49% of cultivated area, plus pasture and woodland pasture and large fields</td>
<td>24,977</td>
<td>0.5</td>
<td>2,430</td>
<td>35,991</td>
<td>138</td>
<td>5</td>
<td>65</td>
<td>674</td>
<td>134.64</td>
<td>12</td>
</tr>
<tr>
<td>Between 30% and 59% of cultivated area, plus pasture and woodland pasture and large fields</td>
<td>25,775</td>
<td>3</td>
<td>889</td>
<td>8,563</td>
<td>63</td>
<td>3</td>
<td>13</td>
<td>559</td>
<td>131.74</td>
<td>11</td>
</tr>
<tr>
<td>Between 30% and 59% of cultivated area, plus pasture and woodland pasture and large fields</td>
<td>2,891</td>
<td>1</td>
<td>155</td>
<td>9,508</td>
<td>40</td>
<td>5</td>
<td>13</td>
<td>581</td>
<td>132.25</td>
<td>11</td>
</tr>
</tbody>
</table>

(continued)
Introduction

Bulgaria’s last agricultural census in 2010 collected information from 370,500 agricultural holdings operating on about 5 million hectares (ha) of utilized agricultural area. Since Bulgaria joined the European Union (EU) in 2007, there has been a continuous decline in the number of farms, while the utilized agricultural area remains stable. Small, mixed (crop and livestock) family-run agricultural holdings have progressively disappeared while field crop holders have increased in size and shares. The more labor and resource-intensive sectors, such as livestock breeding, fruits and vegetables, are in decline. This complex dynamic in the farm and production structure requires corresponding effort to produce high-quality statistics.

Bulgaria’s agricultural statistics comply with the EU’s legislation in terms of objectives, scope, variables and level of precision of estimates. The responsibility for Bulgaria’s production is shared between the Ministry of Agriculture and Food (MAF) and the National Statistics Institute (NSI). MAF, through its Agrostatistics Department, is responsible for agricultural censuses, farm structures and annual production statistics, while NSI produces price statistics and economic accounts for agriculture. MAF uses both a list frame and area frame for its various sample surveys. NSI uses MAF’s data and its own frames.

Methodology for building sampling frame and survey design

Area frame

The area frame sample was built by MAF in 1998 with an objective to produce reliable statistics on land use and land cover, particularly on crop area, at the national and regional NUTS8 2 level. The BANCIK (Bulgarian survey on the

---

8 NUTS = nomenclature des unités territoriales de l’EU
land cover and land use) survey (based on the area frame) was introduced during a period of land restructuring in Bulgaria, when the previously applied methodology for estimating crop area and production was based on the reporting of big state cooperatives, and expert estimates at the municipal level became less adequate, while building and maintaining a list frame for the purposes of agricultural statistics was difficult, expensive and not efficient. Furthermore, BANCIK, representing a direct observation of points, solved another problem related to the difficulties in collecting data directly through face-to-face interviews with farmers. A subsample of BANCIK is used for early estimates of the production of main crops, where the points with certain crops are visited just before the harvest, and crop yields are estimated by crop-cutting methods or by estimates by experienced agronomists (See questionnaire in Annex 1).

The area frame sample is constructed on a network of north–south, east–west oriented straight lines with 6 km of distance between them (Fig. 1).

Figure 1. Bulgaria's districts and municipalities.

Each point of intersection of the network stands for the center of a segment, considered to be random. Some 3 123 square segments with sides of 1 410 m are defined. Each segment contains 36 points that represent the sample of points to be observed in the country. The distance between these points in one segment is 234 m. In total, there are about 111 000 observation points included in the
sample, representing the country’s total area of 111 000 km$^2$; thus, one point represents 1 km$^2$.

Precision is obtained by calculating the coefficient of variation. The total variance is taken into consideration as the sum of the variance between the segments in the sample and the variance between the points of the same segment. The standard deviation is then calculated out of the variance. The precision (coefficient of variation) of the estimation of the area of a given crop $\hat{X}_i$ is calculated as a percentage:

$$CV (\hat{X}_i) = \frac{\sqrt{\text{Var} (\hat{X}_i)}}{\hat{X}_i} \times 100$$

For a given land cover type, the degree of accuracy increases in direct proportion to the area increase and the distribution of the type of land cover observed. For example, because forests cover a considerable part of the territory, the degree of accuracy when calculating the forest area is much higher than that for other less dominant types of land cover. The weight of the points is calculated at the regional level, and takes into account the area of each region.

Initially, topographic maps (scale 1:10 000) and satellite images were used to locate the segments and observation points. Later, satellite images were replaced by orthophotos made for the purposes of the administrative agricultural system. The area frame sample was built and the material to be used was acquired within the framework of cooperation between the Bulgarian and French governments. In preparation of the first BANCIK survey in 1998, more than 5 000 maps were collected, scanned and georeferenced, and 13 000 cartographic documents were used. Dedicated software was developed and installed in regional and central offices of the Agrostatistics Department of MAF.

Preparation of the BANCIK survey began in July 1997 and included building the sampling frame, acquiring maps, GIS packages, and developing administrative procedures). Fieldwork for the first survey was carried out from July to August 1998 by 292 enumerators from MAF and 60 supervisors from NSI. The first results were available in September 1998.

The human resources used for preparing and implementing the first BANCIK survey included:

- 150 working days for project management and supervision and statistical expertise (France)
- 225 working days for technical assistance (France)
• 400 working days for geographical expertise (France)
• 40 working days for IT expertise (France)
• 1 440 working days for the central and regional offices of the MAF (Bulgaria)
• 3 000 working days for the enumerators (Bulgaria)
• 300 working days for the field supervisors (Bulgaria).

The budget necessary for building the sampling frame and selecting the sample, acquiring the materials and carrying out the first year of the BANCIK survey was about EUR 450 000. The project was financed by the United Nations Development Programme, French Ministry of Foreign Affairs, French Embassy in Bulgaria, Bulgarian Ministry of Agriculture, and the European Commission.

**List frame**

The list frame is used for the sample surveys to produce statistics on the farm structure (including agricultural census), on annual production statistics by groups of commodities (e.g. main crops, vegetables, fruits, meat and milk), and on planted and harvested areas of crops and numbers of livestock.

The list frame, which contains a list of all agricultural holdings using the concepts and definitions applied by the European Union, was first built in Bulgaria within the 2003 agricultural census. During the census preparation phase, data on the agricultural activity of households, which were collected during the population census of 2001, were matched with all existing administrative records pertaining to agriculture. The resulting list was cleaned from duplicates and out-of-scope units first (as much as possible), and then from the field during face-to-face interviews during the agricultural census. Between 2003 and 2010, the list was partly updated using data from the sample surveys that were carried out by MAF. New farms, identified through administrative sources or other information, were added to the list and to the relevant samples in a “take-all” farm stratum. The majority of small and medium units of the list frame that were not part of the samples, however, could not be updated. During that period, administrative sources in agriculture were still under development, or were incomplete, or were sometimes incoherent with regard to statistical concepts and definitions.

The list frame was, therefore, fully updated within the 2010 agricultural census. The preliminary list of farms was compiled by the central office of the
Agrostatistics Department, which consolidated information from 12 major sources. Duplicate records were eliminated at the district level by experts at MAF and local authorities, which were well aware of the active holdings in their territory. In total, about 100 people were involved in compiling and updating the list for the 2010 agricultural census.

The time spent for elaboration of the list frame in each of the stages was: three months compiling and cleaning an exhaustive list, and three months collecting field data (agricultural census data collection).

Based on the updated list frame of 2010, new samples were drawn for all thematic annual surveys carried out by MAF, including: harvested area and production of arable crops, harvested area and production of vegetables, area and production of fruits, area and production of grapes, number of livestock. Also included were surveys that were carried out less frequently, such as a farm structure survey, structure of vineyards and structure of orchards.

Depending on the survey scope and precision requirements, the sample size varied from 4% to 8% of the target population of agricultural holdings for a given survey (e.g. crop producers for crop survey, dairy farms for milk survey, total population of holdings for the farm structure survey). The stratification criteria varied according to the survey’s objectives but were usually based on the geographical location, size and specialization of the farms. One take-all stratum includes all very large farms and those with specific production.

**Results obtained and precision of estimates**

**Area frame**

The sample is designed to be representative at the regional (secondary) level of the NUTS classification system. For Bulgaria, there are six regions. If considered necessary, this kind of sample allows the inclusion of new segments in order to obtain reliable results at a lower geographical level. With the current sample, the sampling error of estimate for arable land and main field crops is between 1% and 5%. Precision is not as good for crops that are grown on small areas. Considering the structure of Bulgarian agriculture, these crops include vegetables, tobacco and permanent crops. The table below provides an

---

9 Sources included the 2003 agricultural census and consecutive sample surveys, administrative register of subsidised farms, veterinary register, agricultural producers register, organic farming register, and others.
example of the precision of results for some important agricultural crops and land use categories.

<table>
<thead>
<tr>
<th>Crop/LU category</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>2.24%</td>
</tr>
<tr>
<td>Barley</td>
<td>4.71%</td>
</tr>
<tr>
<td>Maize</td>
<td>3.72%</td>
</tr>
<tr>
<td>Sunflower</td>
<td>4.02%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crop/LU category</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco</td>
<td>8.34%</td>
</tr>
<tr>
<td>Meadows</td>
<td>3.18%</td>
</tr>
<tr>
<td>Fallow land</td>
<td>3.73%</td>
</tr>
<tr>
<td>Vineyards</td>
<td>7.59%</td>
</tr>
</tbody>
</table>

**List frame**

Samples were designed in order to obtain good precision at the regional NUTS 2 level. Although not required by the relevant EU regulation, the precision of estimation is considered to be good if the sampling error is up to 5%, and is considered acceptable if between 5% and 10%.

Examples for standard errors in the farm structure survey of 2013.

<table>
<thead>
<tr>
<th>Crop characteristics</th>
<th>BG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of cereals</td>
<td>0.4</td>
</tr>
<tr>
<td>Area of dried pulses and protein crops</td>
<td>3.0</td>
</tr>
<tr>
<td>Area of potatoes</td>
<td>4.9</td>
</tr>
<tr>
<td>Area of oilseed crops</td>
<td>0.6</td>
</tr>
<tr>
<td>Area of fresh vegetables, melons and strawberries</td>
<td>2.4</td>
</tr>
<tr>
<td>Area of flowers and ornamental plants</td>
<td>7.6</td>
</tr>
<tr>
<td>Area of pasture and meadow, excluding rough grazing</td>
<td>1.8</td>
</tr>
<tr>
<td>Area of fruit and berry plantations</td>
<td>2.2</td>
</tr>
<tr>
<td>Area of vineyards</td>
<td>1.7</td>
</tr>
</tbody>
</table>

**Livestock characteristics**

<table>
<thead>
<tr>
<th>Livestock characteristics</th>
<th>BG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dairy cows</td>
<td>1.1</td>
</tr>
<tr>
<td>Other cows</td>
<td>3.7</td>
</tr>
<tr>
<td>Other bovine animals</td>
<td>1.3</td>
</tr>
<tr>
<td>Breeding sows</td>
<td>1.3</td>
</tr>
<tr>
<td>Other pigs</td>
<td>0.5</td>
</tr>
<tr>
<td>Sheep</td>
<td>2.2</td>
</tr>
<tr>
<td>Goats</td>
<td>2.8</td>
</tr>
<tr>
<td>Poultry</td>
<td>1.1</td>
</tr>
</tbody>
</table>
Up until 2010, the area frame survey (BANCIK) was used as a reference for arable land and main crops area. It should be noted that for these land-use categories, the difference between BANCIK and the agricultural census was minimal, around 1%. Data were also cross-checked with administrative sources (mainly administrative agricultural system). The yield and production of annual crops were then calculated with a specialized annual survey on main crops.

**Resources mobilized and other costs**

**Area frame**

After building the area frame sample and acquiring the survey materials, the costs for the annual data collection decreased significantly. The same methodology is used and the same points are visited every year. Satellite images initially used were replaced after 2007 by orthophotos that were developed for administrative agricultural purposes and acquired for the purposes of the BANCIK survey for no additional costs. The estimated annual cost is approximately EUR 30 000, including data collection, data control and data processing, transport costs for fieldwork, and office expenditures for questionnaires and publications.

**List frame**

Once the list frame is built, it must be regularly updated using statistical sources (sample surveys carried out during the years) or administrative sources (agricultural registers kept at MAF or relevant agencies in the agricultural sector). The use of administrative sources, however, does not require additional costs because access to administrative sources for statistical purposes is free of charge.

The annual cost for thematic production surveys represents the salaries of about 100 experts from MAF’s Agrostatistics Department and regional directorates. For farm structure surveys and censuses additional funds are required for extra enumerators and supervisors.

With regard to survey design and organization, one person at the central level is designated as being responsible for each survey. The responsible person organizes the entire survey process. He/she designs the questionnaire and guidelines for enumerators; trains district staff; develops the data entry application; controls, processes and validates the data; and prepares the data dissemination. One person at the central level is responsible for the design of the sample for each survey supported by the expert in charge of the survey.
At the district level, staff at the Agrostatistics Department are responsible for identifying new farms to be added to the sampling frame, organization of the data collection process, selection and training of enumerators, control of fieldwork, and data entry.
Survey data collection

Paper and electronic questionnaires (for Internet and email data collection) are
developed and used for data collection for list frame surveys. Survey documents
are developed by experts at the Agrostatistics Department. The area frame
(direct observation) survey uses paper questionnaires.

Fieldwork is organized and managed by district staff of the Agrostatistics
Department, who supervise the enumerators. There is a pool of experienced
experts at the district and municipal levels who can be used as enumerators, and
are trained for each annual survey. In case of larger surveys (such as farm
structure surveys and censuses), additional enumerators and supervisors are
hired.

Enumerators are trained (for one to two days, depending on the survey) in
conducting the questionnaires and the guidelines for the respective survey, the
list of farms to be visited by them, and are given supporting material relevant to
the survey (e.g. maps, catalogues, list of crop varieties). The work of
enumerators is controlled in the field and by phone. Up to 5% of the
questionnaires may be subject to control, meaning that 5% of the filled-in
questionnaires are chosen and the farms are interviewed again by supervisors,
for control purposes; or, the farms are at least contacted and asked some
questions on the basis of the questionnaires filled in by the enumerators. The
purpose is to control the enumerators, to ensure that they actually visited the
farm and interviewed the right person.

Data entry for paper questionnaires is done at the district level by operators who
are selected, trained and supervised by district staff of the Agrostatistics
Department. After validating, district data are sent to the central level of the
Agrostatistics Department for further control and validation, processing and
dissemination.

Issues and challenges encountered and how they were addressed

List frame

A list frame can be fully updated only with an agricultural census. Its regular
update in between two censuses (usually 10 years) depends to a great extent on
the quality of the existing administrative sources, their objectives and scope,
basic unit of registration, characteristics collected, definitions used, and
completeness. The existence of a large number of very small family farms in the
structure of Bulgarian agriculture just after the land restructuring process in the 1990s limited the use of an updated list frame. These small farms are usually not registered in any of the registers and, therefore, their characteristics cannot be updated on a regular basis. Another issue with family farms (even larger ones) is that one farm could be registered in different administrative sources with different family members. This makes it difficult to match records from different sources.

Because of these constraints, in the period 2003–2010 Bulgaria used the same sample for the annual surveys as a panel, adding new farms in a take-all stratum. This approach cannot ensure that all new farms (especially small farms) that appeared after the agricultural census in 2003 were covered and, thus, there could be undercoverage. In 2010, when updating the list frame for the 2010 agricultural census, there were about 10 000 new farms identified during the fieldwork (i.e. 2.7% of unit undercoverage of the sampling frame used before 2010, although their impact in terms of utilized agricultural area was small).

The evolution of small farms up until 2010 depended more on general economic development than on the farms themselves. Therefore, it was decided that for the purposes of annual livestock surveys, two subpopulations be defined: professional farms and family farms. For family farms, a village survey was carried out. Villages were stratified based on their (human) population, and not on their population of livestock. A sample of about 250 villages was selected out of 5 200. These villages were enumerated door-to-door in order to collect the annual (or more frequent, when necessary) data about livestock every year. As an example, in 2004 the village survey represented about 80% of the livestock breeding farms and 32% of the cattle breeding farms, 29% of pig farms, 45% of sheep farms, 74% of goat farms, and 40% of chicken farms. This share decreased progressively each year thereafter. The village survey represented a cost-effective tool for producing good quality statistics on the number of livestock and livestock products in a situation where there was a large number of very small and dynamic farms with significant shares in the total livestock breeding activity, and a lack of reliable sources for regular updating of the list frame. In 2010, after an analysis of the structure of the livestock sector showed that breeding activity had been restructured and was predominantly carried out by professional farms, the village survey was discontinued thereafter.
Conclusion, lessons learned and remaining issues

Area frame

The main advantage of using an area frame sample survey (BANCIK) was that it provided reliable statistical data in a period when other methods were considered either less reliable or less feasible. Building and maintaining a list frame of the farm population, which was going through dynamic structural changes in the 1990s, required significant human and financial resources. The first agricultural census after the land restitution was carried out in 2003. Even after that, profound changes in farm structures continued. Thus, the area frame sample survey (BANCIK) provided reliable statistical data on agricultural area at the national and regional level. Being a sample survey, a certain level of precision can be calculated for its results compared with previously used expert estimates. BANCIK collected data through direct observation of points, thus eliminating the subjectivity of farmers’ declarations (which can often be wrong or false). In order to minimize the observation error, the same points are observed every year. Implemented for more than 15 years now, BANCIK is a powerful tool that not only estimates the area and sequence of crops, but also provides information on change in land cover and land use as a result of the structural changes that took place at the beginning of the 21st century, and the implementation of measures of the Common Agricultural Policy (CAP) of the European Union in Bulgaria since 2007.

The disadvantage of this method is that it is less representative for small areas and for crops that are usually grown on small farms such as tobacco, vegetables, orchards and vineyards. It can also generate an error of classification of the land use for categories whose use is difficult to distinguish by direct observation (e.g. permanent grassland, recently abandoned agricultural area, and fallow land).

With the development of the administrative agricultural system, after 2010 the use of administrative sources (registers) gradually replaced the use of BANCIK for agricultural statistics. Now it is used for producing land use and land cover and agri-environmental statistics, including statistics on area on which crop rotation is applied and prior crops and for early estimates for wheat and barley.

List frame

In the period after 2010, agricultural administrative registers improved considerably in terms of relevance and coherence with statistical scope and definitions. On the other hand, the share of very small family farms decreased significantly after 2003, and this trend continued after 2010. Therefore, the use of administrative sources for updating the list frame, and even for directly
producing some main agricultural statistics indicators, is becoming more and more frequent. MAF’s Agrostatistics Department continually analyses these sources and evaluates their possible statistical use. In the years to come, it is planned to replace some of the regular sample surveys with administrative sources. This can be done for open field crops (mainly cereals and oleaginous crops), which are usually grown by large farms. However, the share of small family farms remains significant in the production of fruits and vegetables, tobacco and aromatic plants where sample surveys will continue to produce better statistics in the coming years.
References


Ministry of Agriculture and Food, Agrostatistics Department. 2014. Land-cover and Land-use Survey BANCIK – Final Results for 2014. Results and Analyses, issue # 276, October 2014.

Ministry of Agriculture and Food, Agrostatistics Department. 2015. Land-cover and Land-use Survey BANCIK – Final Results for 2015. Results and Analyses, issue #297, October 2015.

For more information, see:

ANNEX 1

**QUESTIONNAIRE for yield forecast for the main crops - wheat and barley**

<table>
<thead>
<tr>
<th>No of the point</th>
<th>Code</th>
<th>Crop</th>
<th>Variety</th>
<th>Density of crops</th>
<th>Word density</th>
<th>Decays and parasites</th>
<th>Yield</th>
<th>No of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>11</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>12</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>13</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>16</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>17</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>19</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>21</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>24</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>25</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>26</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>27</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>28</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>29</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
<tr>
<td>30</td>
<td>23</td>
<td>24</td>
<td>______</td>
<td>______</td>
<td>2</td>
<td>______</td>
<td>3</td>
<td>______</td>
</tr>
</tbody>
</table>

**Use “X” to mark the correct answer**

<table>
<thead>
<tr>
<th>Word density</th>
<th>1 low</th>
<th>2 average</th>
<th>3 high</th>
<th>Decays and parasites</th>
<th>Yes</th>
<th>0</th>
</tr>
</thead>
</table>

42
<table>
<thead>
<tr>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>Aglica</td>
<td>41.84 gr</td>
<td>176</td>
<td>Zlatina</td>
<td>25.3±5.1 gr</td>
</tr>
<tr>
<td>186</td>
<td>Adria</td>
<td>30.5±2.3 gr</td>
<td>180</td>
<td>Zora</td>
<td>35.8±3.2 gr</td>
</tr>
<tr>
<td>187</td>
<td>Adria</td>
<td>26.5±1.6 gr</td>
<td>189</td>
<td>Zora</td>
<td>36.2±3.5 gr</td>
</tr>
<tr>
<td>196</td>
<td>Adria</td>
<td>20.0±1.1 gr</td>
<td>198</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>200</td>
<td>Adria</td>
<td>18.0±1.1 gr</td>
<td>199</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>201</td>
<td>Adria</td>
<td>16.8±0.9 gr</td>
<td>200</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>202</td>
<td>Adria</td>
<td>13.0±0.9 gr</td>
<td>201</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>11.8±1.1 gr</td>
<td>202</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>9.8±0.9 gr</td>
<td>203</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>8.8±0.9 gr</td>
<td>204</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>7.8±0.9 gr</td>
<td>205</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>6.8±0.9 gr</td>
<td>206</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>5.8±0.9 gr</td>
<td>207</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>4.8±0.9 gr</td>
<td>208</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>3.8±0.9 gr</td>
<td>209</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>2.8±0.9 gr</td>
<td>210</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>1.8±0.9 gr</td>
<td>211</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>0.8±0.9 gr</td>
<td>212</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
<tr>
<td>203</td>
<td>Adria</td>
<td>-1.8±0.9 gr</td>
<td>213</td>
<td>Zora</td>
<td>36.2±3.2 gr</td>
</tr>
</tbody>
</table>

**Varieties of Winter Wheat**

<table>
<thead>
<tr>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>207</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>208</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>209</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>210</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>211</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>212</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>213</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
</tbody>
</table>

**Varieties of Winter Barley**

<table>
<thead>
<tr>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>207</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>208</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>209</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>210</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>211</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>212</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>213</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
</tbody>
</table>

**Varieties of Winter Oats**

<table>
<thead>
<tr>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>207</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>208</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>209</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>210</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>211</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>212</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>213</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
</tbody>
</table>

**Varieties of Winter Rye**

<table>
<thead>
<tr>
<th>Code</th>
<th>Variety</th>
<th>Weight of 1000 grains</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>207</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>208</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>209</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>210</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>211</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>212</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
<tr>
<td>213</td>
<td>Adria</td>
<td>25.8±2.3 gr</td>
</tr>
</tbody>
</table>
Summary

Since 2003, China’s National Bureau of Statistics (NBS) has been studying and applying remote sensing to crop areas and yield measurements. In recent years, the methodology and data resources have made significant progress. NBS started from the remote sensing method of research, and has progressed to the operational stage of remote sensing to measure the area of grain crops and important economic crop areas. This development is the outcome of combined actions of several factors: 1) data sources for remote sensing application have effectively been improved; 2) the quality of remote sensing images and auxiliary data has significantly improved; 3) the system of an agricultural statistical survey that is based on remote sensing has been preliminarily established. (Many provinces have developed a crops area frame survey and a current remote sensing measurement survey for use in determining the planting area of major crops.); and 4) software and hardware for remote sensing application has improved. As the crops area frame survey and the main crop planting area of remote sensing measurement survey became operational, the space-air-ground integrated agricultural statistical survey framework was well developed, and the agricultural statistics system was built for the remote sensing operational survey.

Introduction and background

The agricultural statistical survey conducted by China’s National Bureau of Statistics (NBS) had wide coverage, and relied on a team of highly qualified professional staff with expertise in conducting socioeconomic surveys in agriculture and rural areas across the country. The team included staff from NBS branches in nearly 1000 counties (which constitutes one-third of the total number of counties in China), which when combined, totaled nearly 20 000 full-time staff and a large number of assistant interviewers to manage and conduct the agricultural statistical surveys.

The history of agricultural sample surveys at NBS began several decades ago with the rural sample surveys of 1984, which included an agricultural production survey (area, yield and production), a rural household survey, and a rural socioeconomic survey. The major sampling method used then was multistage
sampling, from province to county, township, village, farmers and farmers’ land. In each county sampled at that time, a county-level branch office was set up and these county branches were maintained afterwards. In 1990 and 1995, two rounds of sample rotations were conducted. The basic pattern of these two sample rotations was to keep sample counties almost fixed, and only selected lower sample units within these counties were changed.

The sampling design used currently is still multistage sampling, although county sampling is no longer done. In 2000 and 2005, sample rotations were conducted in nearly the same way although multivariate probability proportional to size sampling schemes were used for cost-efficiency. In 2000, rural sampling surveys had been enriched, including the yield survey, rural household survey, rural economic survey, and extended to the main crops (rice, wheat, corn, cotton planting area and yield surveys), the major livestock survey, farmers’ assets survey, agricultural product prices and intermediate consumption surveys. The basic idea of the sample design is two-phase, treating all fixed national sample counties as the target population, and then sampling villages within counties and units within the villages.

A more recent analysis of the current Chinese agricultural statistics system, however, highlighted certain weaknesses and issues that were considered to negatively affect the accuracy of agricultural production statistics. These issues included the rapid development of China’s economy and society, especially as a new rural construction strategy concerning the transfer of rural land operational rights and agricultural market fluctuations were being implemented. The system was no longer adequate due to the following:

- The fact that the system is based on sampling surveys for major crop production statistics and on comprehensive reporting from the administrative system for other crop production statistics (e.g. oil seeds, sugarcanes, fruits, vegetables and other crops) raises the issue of how to integrate the results from the sample estimation of overall crop area and major crops area with comprehensive reporting results.

- China’s agriculture is rapidly changing with the growing flux of rural farmers away from rural areas; in these cases, holdings cannot be easily surveyed, which could lead to estimation bias and poor sample stability.

- Survey tools for current agricultural surveys mainly consist of traditional tools, such as compasses and rope for area measurements, while modern tools, such as a GPS (global positioning system) or PDA (personal data assistant), are rarely used. The means for information transmission is also traditional, consisting mainly of paper and pen. Survey methods do
not match survey tasks, and the lack of technical means for effectively supervising and managing data often-times means the results include human errors.

The degree of impact of these constraints on agricultural statistics, such as those for crop production, is considered to be very high, and it was determined that the existing survey system must be reformed and innovated; hence, the decision to build a new survey framework.

The new system is based on geospatial agricultural surveys, with mainly area frame surveys, complemented by current satellite or remote sensing measurements and integrated space-air-ground surveys and monitoring methods. The system will be set up by developing the full-coverage area frames, based on cultivated land data, agricultural census data, and satellite imagery. The spatial sampling method will be used to select georeferenced villages and cropland segments, and taking cropland and crops as counting units, crop planting area, yield and related surveys will be designed and conducted. GPS, PDA and other new technological tools will be used to improve the efficiency and data quality of the survey. The use of an area frame survey and remote sensing are considered to be major reforms in agricultural production surveys in China. Major innovations include:

1. **Changing the sampling methods and the survey methods.** The crop planting area survey will be changed from “farmer-based” to “land-based”, and from “subjective” to “objective”. Based on the area sampling frame of cultivated land, crop planting area surveys for grain, cotton, oil and sugar crops will be developed.

2. **Applying new technology to the crop production survey (mainly grain and cotton).** In the crop planting area survey, new advancements GIS and GPS (global positioning system) will be widely applied, and PDAs will be used to collect data. This should result in an improvement in work efficiency, measurement accuracy, the quality of supervision, and inspection.

3. **Reducing the risks of declaration bias to the statistical data.** In an area frame survey and remote sensing measurement, land plots will be considered as counting units, which cannot be influenced by subjective effects and, thus, have complete repeatability, and can also be checked. This will effectively overcome artificial interference and ensure the authenticity of the survey because data will come from direct measurements during field surveys and investigative work. This method
will also reduce the possible interference of the farmer, and guarantee the objectivity of the survey data, and will also facilitate tracking the data quality assessment of data.

In recent years — with the support of the national "863" project\textsuperscript{10} of science and technology, national satellite application industrialization project support — the application of spatial information technology at the provincial level was tested. The test, demonstration and verification were conducted in Beijing, Heilongjiang, Liaoning, Jilin, Jiangsu, Zhejiang, Anhui, Shandong, Henan, Hubei, Guangdong, Ningxia, Xinjiang and some other regions. The results show that the estimation precision for target crops (rice, wheat, corn, cotton) reaches the operational requirement. Results produced at the county and prefecture levels provide a solid basis for policy-making, implementation and monitoring. Satellite navigation and electronic data capture in the field will allow real-time monitoring of field survey data, and provide an effective means to improve the survey’s quality. Upgrades to the agricultural statistical survey system (in the application system), survey tools and methodology were realized, as shown by preliminary results.

**Methodology for building a sampling frame and designing a survey**

**Current remote sensing survey**

In order to efficiently and accurately determine the planted area of major grain crops in major grain-producing provinces, NBS conducted a crop area frame survey and the current remote sensing measurement in tandem. The area frame survey was designed to estimate the total planted area of crops, the planted area of grain, and the proportion of main crops. Remote sensing is for estimating the key crops (part of the main food crops and economic crops, including wheat, corn, soybean and cotton) within a planted area. The two types of surveys will complement one another and confirm results in order to obtain objective and accurate results.

Remote sensing measurements will comprehensively utilize multi-source, multi-phase remote sensing data, and combined with land-use classification data (according to the different development stages of the crops on the remote sensing

\textsuperscript{10} Chinese National High Technology Research and Development Program (863 Program)
image spectral differences), extract the target area of farming land mainly with the artificial visual interpretation method. Farm land refers to the data of land-use classification of cultivated land within the scope, by removing specific changes in land use, such as buildings and water, as well as any temporary change in land use within the survey year for areas where no crops will be grown, such as nursery, mulberry field, and tea. On the basis of unsupervised classification, supervised classification and automatic change detection, a variety of methods were used to identify the spatial distribution range of crops. At the same time, sampling expansion estimation, matrix correction, and methods such as regression model were used to estimate the planted area of crops. Accordingly, statistics on the planted area of main crop provide scientific and objective reference data. Remote sensing measurement work is described below.

**Cropland extraction**

On the basis of land-use classification data and county-level administrative boundaries, GIS software was used to organize and carry out the land-for-farming extraction operation. Based on the high and moderate resolution remote sensing data, land-use classification data of cultivated land was checked and updated by direct observation. Based on high resolution remote sensing data, data updates were reviewed to ensure quality and accuracy of the crop land extraction.
Remote sensing measurement of the area of planted for rice

The spatial distribution of rice to extract

According to the phenological calendar for early rice, middle rice and double-season late rice (remote sensing data can distinguish between two types of late rice), and combined with the latest high and medium resolution remote sensing image and the use of remote sensing professional software for unsupervised classification, supervised classification and automatic change detection, accurate identification of early rice, late rice, middle-season rice and double season late rice spatial distribution can be carried out.

Area estimation

Based on the spatial distribution of early rice, middle rice, one-season late rice, double-season late rice and cropland for the target region, the remote sensing area estimation method based on sampling, an accurate estimation of the planting area of target crops was carried out. At the same time, an estimator was calculated, based on the sampling design and the method based on the model of the early rice planting area. Precision and efficiency were analyzed accordingly.

Remote sensing measurement for other crops

There are many crop identification methods. According to phenological information, the use of different crops in different key phenological periods of different phenological characteristics is more practical and has a higher level of precision.

Data acquisition and processing

According to existing data sources, and based on the phenological phase for remote sensing crop identification, data acquisition should follow the following principles:

- Ensure that key phenological phase image is complete

At certain times and for key phenological phase images, remote sensing images can differentiate target crops with other features because they have an obvious spectrum. With data acquisition, the first step is to obtain images during key periods and at the maximum extension, especially at seed time and harvest.
• Use high spatial resolution as much as possible

Priority is given to selection of high spatial resolution image sensor data, such as resource 3 and resource 1-02c, which are high resolution series satellite data. The second choice should be the medium resolution image data, such as environmental satellite data, Landsat TM data, and others. If there is no high resolution image data are available, then medium resolution image data with the same critical period should be used. The remote sensing images used for crop classification should be multiple spectral images. Panchromatic image should be fused with multispectral images of similar sensor.

• Image quality is reliable

The spectrum should have a uniform color, medium contrast, no obvious “noise”, spots or bad lines. Cultivated land plots in key areas should have less cloud and snow cover. Remote sensing data from different sources must be standardized, including image cutting, rectification and correction. When compared with the reference images, control points in mountainous areas are residually controlled within two to three pixels, plains are controlled within one pixel.

**Crop identification and estimation**

• Defining major crops

Major crops should be defined within each county. All crop planting areas from the previous year are sorted, and crop areas in excess of 10% of the total planting area of crops are treated as major crops. County data from the previous year can be a reference for the comprehensive data of crop planting area.

• Choice of identification method

There are many crop identification methods, each with their own advantages and disadvantages. According to practical experience accumulated over the years, it is necessary to use different displays in which different crops in key phenological periods (e.g. seeding, growth, harvest) are separated from other features on remote sensing images, to establish judgment logic for identification. This identification process needs to obtain accurate phenological data, establish the corresponding spectral library, and select a scientifically feasible and highly accurate classification identification method.
Phenological data acquisition and analysis

Phenological data acquisition should be done for each county. For each major crop, information on seed time, green period, and growth (e.g. the main growth period at harvest time) should be gathered, and some main planting area information should also be obtained. At the same time, information on the main county non-farm crops, such as planting seedling, traditional Chinese medicine, aquaculture, and others, and the corresponding typical areas information, should be obtained. The analysis of corresponding phenological information can be done on the main crop varieties to help with crop identification.

Data aggregation

Because stratification was done at the time of sampling according to the village, the identification results of major crops must be carried out in accordance with the administrative village. Administrative village data are obtained from the second land survey data; therefore, more accurate control and validation work must be done.

Area estimation

Based on the classification results of remote sensing data and crop land data, the remote sensing area estimation method for target regions is based on sampling, which gives a relatively accurate estimation of the planting area of target crops. At the same time, estimation based on the sampling design and estimation based on the model were analyzed for accuracy and efficiency evaluation.

Crop area frame survey in major provinces

When choosing the agricultural statistics method based on remote sensing, data availability and data quality are very important. If the sky is cloud-free or there are few clouds, then medium or high resolution satellite data cannot determine the effective coverage during the crop’s key phenological period, and current remote sensing measurements will be difficult to be implemented. This kind of situation occurs frequently in some complex conditions, such as in southern China. But the area frame sample survey, which mainly depends on the probability of survey sampling design, can be designed and implemented by using relatively stable, objective reporting units, and effectively sampling from an area frame or multiple frames that mix an area frame and a list frame to realize the integration of all kinds of agricultural statistical surveys. The new survey framework can not only meet the demand of multipurpose agricultural statistics, but can also develop a master sampling frame for agricultural statistical surveys.
and build up the data collection basis for current agricultural surveys and agricultural censuses.

According to the global strategy to improve agricultural and rural statistics proposed by the international community, the implementation of area frame sampling and a master sampling frame will provide survey framework to collect information from all aspects of agriculture and rural development, including households, farms, or agricultural production and operation units, agricultural production and operation facilities and administrative areas. It also provides the connection between the census framework and the land use. Finally an integrated survey framework for agricultural statistics will be established. On the basis of the integration of agricultural statistical investigation framework, timely, consistent agricultural statistics can be produced and provided.

With reference to the successful experience of the international community, NBS initiated a series of pilot surveys and researches to promote the reform of agricultural statistics. In accordance with the reality of our country, crops area frame sampling survey methodology was developed.

The design of the crops area frame survey fully considers the more and more available satellite images, geographic information systems and handheld navigation and positioning devices as an important means of building an operational survey system. These means have now become the most practical way for the area frame survey to producing annual and seasonal data.

**Objectives of a crops area frame survey**

Periodic (annual and seasonal), wide range (national and provincial), multi-objective probability sampling of crops survey can get timely and reliable basic data. In China, major target indicators of crops survey include: Farmer’s planting intention, crops planting area, crop harvest area, crop yield (forecast and actual), total crop production, the social and economic characteristics of actual cropping operator and villages they reside.

For the variables of interest and their accuracy requirements, the project objective was defined clearly at the beginning, because this requirement is highly related with the overall sampling design, especially with the type of estimator, questionnaire design and data collection process. Concerning timeliness requirements, survey data should be completed within a month of collection, summary and release.
Survey sampling design

A stratified two-stage sampling method was adopted, in which the primary sampling unit consisted of administrative villages, and the final sampling unit was the segment composed of cultivated plots (e.g. a natural field or a field combination, with a clearly identifiable physical boundary). The first stage of compiling a sampling frame was to stratify the administrative villages within the target population, by using crop density (crop area for cultivated land, using data from census enumeration or remote sensing classification). All these sampling units are geographical areas that may have clearly identifiable natural boundaries or administrative boundaries.

The survey involves annual and seasonal field data collection work, which demands that surveyors complete a questionnaire for each sample unit area. Surveyors access the actual cropping operator or local village officers who are familiar with the operators in order to collect information on sample segments. In addition, to filling in the questionnaires with holders, data collection also involves agricultural area recognition and measurement. For each sample segment, surveyors use several kinds of maps, including image maps and vector maps. For each field in the sample segment, surveyors need to draw the physical boundary of the field on the map. Surveyors also need to verify the crop area planted and other land uses, and any information provided by the operator.

Major survey indicators

A crops survey mainly collects and estimates indicators about crop production, which include the planting area, yield, and total output of main crops such as rice, wheat, corn and cotton. In addition, information on the planting area of other minor crops, yields and total output will also be collected and estimated.

Development of the sampling frame

The compilation of the survey sampling frame will use several types of information. The first source is the second national agricultural census data, including the enumeration of planted area for crops from the 2006 census. The scope of these data is in line with the basic number of cultivated land holdings registered in farmers’ land contract certificates. There are, however, some aging problems with these kinds of data. The second source of information is the second national land-use survey. Although these are geospatial data on cropland, they lack information about crops. The third source of information is remote sensing imagery, which can be processed to obtain more recent agricultural spatial information, such as the planting of main crops. But with the limited crop classification accuracy, the estimation accuracy of planted land area is not
enough. When merging and calibrating these data, careful considerations should be taken with the specific situation of the local crops. Data from several sources of information will be synthesized so that they complement each other with regard to the county or the province. These data will be used to choose proper PPS size measures and stratification variables, and to compile the sampling frame.

**Sampling unit**

The most basic unit of the crops area frame survey is cultivated land, which can be observed directly, and many direct observation field surveys were done. The segment size, however, cannot be set too large. A suitable segment size should consider the sampling variance comparison of different segment sizes, the proportion of non-zero crop area value, survey cost, surveyors’ daily workload, and the accessibility of segment physical boundaries on the images. After considering the above factors, we selected segments of 2 hectares (about 30 mu) in Jiangsu and Hubei, and a segment of 5 hectares (75 acres) for Jilin, Henan and Liaoning. Considering the current situation of the sampling frame data source, we do not sample segments directly, but rather use multistage sampling so that the frame is compiled stage by stage.

Within the spatial area of the target population, the administrative boundaries are used to split village space. Information on the planted area of crops come mainly from the second national agricultural census enumeration data, combined with remote sensing measures, and the second national land-use survey data for calibration and matching. Finally, the sampling frame of administrative villages as the primary sampling unit was compiled.

Within the spatial area of the sampled primary sampling unit, the cultivated land is split or combined with the natural boundary to create segments; therefore, the frame of cultivated land segment for sampling was compiled.

**Sample selection**

*The sample for crop planting area*

A two-stage sampling method was adopted to select samples for the crop planting area. Specific steps were as follows: In the first stage, the stratified probability proportional to cultivated land method was used to select primary sampling units, which formed sample sites. The precision requirements for major crops (total planting area of crops, grain planting area, major grain crops planting area) are controlled within 5% CV (the relative error coefficient, or coefficient of variation). With this limit, the sample size was determined by
trade-offs among workload, cost and accuracy. In the second stage, sample segments were selected within the sample village by the simple random sampling method. Specific steps were as follows: Within the villages, all cultivated land segments were created and ordered in serpentine order. A random number table was used to generate random numbers, and then five segments were selected as the sample.

*The sample for crop yield*

A two-phase sampling method was adopted. A crop yield sample was selected from the sample of the crop planting area.

**Data collection**

*Screening survey*

Sample segments of cultivated land were screened at the beginning of the survey. The data collected include the sample location, the long-term land use and area covered, splitting of physical fields, and number of operators. The basic information survey of the villages where sample segments were located was conducted in September every year, mainly including information on the population and labour force, the condition of cultivated land, crops, and other factors. At the end of August, the major crop production is forecasted, and the forecast method is based on the current major crops condition, and is compared with the previous year. In order to determine the major crop production of the current year, the crop yield was forecasted first. For the crops that have been harvested, actual productions will be used.

*Seasonal survey*

Seasonal surveys were conducted and reported according to crop growing season. Planting intention survey of farmer will be conducted three times, planting area survey three times in winter, spring and summer, planting period, and crop yield three times in summer early rice, autumn harvest period.

**Data processing**

*The data capture of screening information*

County survey branches are responsible for data capture of paper questionnaires by using dedicated survey data processing software. After the data capture, the data will be audited and transmitted to the central office through a dedicated network.
GPS devices for collecting information

County survey branch staff should export all data within the GPS devices to files with a predefined format. Then, all exported files should be check and packed into compressed files. Finally, the compressed data file should be transmitted to the central office through a dedicated network.

Archiving the registration forms of the screening survey

The team responsible for the sample unit of cultivated land area registration survey keeps the original table by county, sample copies of cultivated land registration forms, sample unit area and unit area. An auxiliary investigator is responsible for warehousing the lists of names of operators by village and per season, as the basis of the survey.

Data assurance, software and IT infrastructure

Data assurance system

Satellite imagery assurance system

The satellite imagery is composed mainly of domestic satellite data, and complemented by some foreign and commercial satellite data. Domestic satellite remote sensing has developed rapidly in recent years, and the support capability of satellite remote sensing data has increased, thus reducing the cost of remote sensing. Domestic satellite data are constantly available and include high-resolution series satellites, resources series satellites, and environments series satellites. Through the efforts of recent years, we have formed a good data cooperation mechanism, through line ministries and commercial partnerships. Line ministries include the Ministry of Land Resources, the National Administration of Surveying, Mapping and Geo-information, the National Fundamental Geographic Information Center, and others. Commercial partnerships include Oriental Roadnear, Peacemap, China Science Geo-do, and others.

Field survey data acquisition

Data acquisition consists of:

1. Area measurement and data acquisition with mobile electronic devices enhanced with GPS.
2. Rahman–Pinty–Verstraete (RPV) data acquisition.
3. Photogrammetry, such as small aircraft, unmanned aerial vehicles (UAV or drones) and airship, is developing very fast with costs being largely reduced. The application for rapid monitoring has become relatively mature, and a social services system has been gradually formed. It can be easier to utilize through buying service.

4. The agronomy parameter acquisition toolbox.

5. Integration of vehicle-based data acquisition system.

Vehicle-based data acquisition system is a kind of powerful, integrated field data collection system, which includes the function of a data center, survey center, monitoring center, emergency center and transportation center. It can be used for crop planting area surveys, crop condition monitoring, and crop-cutting yield surveys. It can receive real-time survey tasks, and navigate to the survey site quickly. A handheld GPS is used to measure the crop planting area. Vehicle-mounted UAV can capture landscape photos on a large scale. A vehicle-based data processing system can process images, videos, other types of unstructured data and upload the results to a survey center quickly.

*Baseline images, basic geoinformation and other auxiliary data*

1. Total coverage remote sensing baseline images

Through research projects in recent years, the satellite application demonstration projects and industrialization projects, a suite of total coverage, remote sensing baseline image datasets has been gradually formed, including multiphase, complete coverage, high-resolution remote sensing image datasets.

2. Basic geoinformation

Based on the second land-use survey, geoinformation datasets by province, county and village, land plots were constructed.

3. Other data

This mainly includes the spatialized statistical data, climate data, phenological data and hydrological data.

With the operational application of remote sensing agricultural statistics expanding, the standard, norm and technology system for the agricultural statistical sampling frame construction has been basically developed. By taking advantage of stable and high quality of domestic satellite remote sensing data, and the implementation of the three-year plan, the construction of the geospatial
data framework of agricultural statistics for 13 grain-producing provinces and/or 1200 grain-producing counties can be realized. This construction plan includes: an agricultural statistics spatial sampling frame, basic geoinformation, high-resolution remote sensing baseline image datasets, and agricultural statistical thematic datasets, which will provide a solid data foundation for geospatial agricultural statistical surveys in the future.

**Operational remote sensing system for agricultural statistics**

A suite of agricultural statistics for an operational remote sensing system was developed with the aim of supporting the agricultural statistical survey through high-resolution satellite application project\(^\text{11}\), and domestic satellite industrialization project\(^\text{12}\), and by integrating related remote sensing component of business model for agricultural statistics. The remote sensing system includes: a business management subsystem, multivariate data management subsystem, agricultural production statistics subsystem, and a county-level field survey tasks transmission subsystem. Among these subsystems, the business management subsystem, multivariate data management subsystem, and agricultural production subsystem deal with data management and create the survey tasks and perform data production, which is the center system. According to business requirements, these subsystems can be deployed at the national or provincial level. The function of each subsystem is described below.

**Survey conducting subsystem**

Based on the business flow of an agricultural statistics rapid survey, the core functions of a business management subsystem include user and role setting, authentication, survey task scheduling, task monitoring, system resources management, and others.

**Data management subsystem**

For the requirements of a unified and standard storage system and the efficient management of multidimensional data, data are categorized as agricultural statistical sampling frame data, remote sensing data, sample unit basic data, and data products. Defining the interfaces between the integrated service platform for an agricultural statistics rapid survey and a data sources assurance system of domestic satellite remote sensing, multidimensional remote sensing data needed for agricultural statistics production can be gradually acquired. By providing a

---

\(^{11}\) [https://www.globalsecurity.org/space/world/china/cheos.htm](https://www.globalsecurity.org/space/world/china/cheos.htm)

\(^{12}\) [https://www.google.rw/search?q=china+domestic+satellite+industrialization+project&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKEwjm8saSz4bYAhVEvhQKHSohB8gQsAQIoQg&biw=1280&bih=574](https://www.google.rw/search?q=china+domestic+satellite+industrialization+project&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKEwjm8saSz4bYAhVEvhQKHSohB8gQsAQIoQg&biw=1280&bih=574)
metadata registry for different types of data, the management of multivariate data are highly effective. By using a distributed multiple data storage model, the equilibrium load of the data server, and the speed and response ability of data access will increase. By adding, deleting, and modifying, the effective access and update of the multivariate data can be realized.

**Agricultural statistical production subsystem**

An agricultural statistical production system is a rapid core business subsystem in the integrated services platform of an agricultural statistics rapid survey, which includes the rapid calibration of satellite images, sampling frame maintenance and update, developing a sampling scheme and sample selection, area estimation, condition monitoring, production forecast and agricultural disaster monitoring, compiled and issued with survey tasks and survey data processing.

**Data query, analysis and release system**

Two of the objectives and tasks of spatial agricultural surveys include research and construction of a spatial visualization survey data query and browser system, and setting up standard and norm for data productions and services. According to different user requirements, GIS and multimedia technology, will be used to disseminate information in more convenient and clear way (e.g. using mobile terminals, WebGIS).

Based on the statistical analysis of existing products, and combined with spatial analysis technology, this system will provide user-oriented data visualization, interactive analysis, basic statistics analysis, multivariate statistical analysis, time series analysis and exploratory data analysis methods, thereby improving the efficiency and accuracy of the statistical analysis, and meeting the diversified and multi-level user demand for agricultural statistics analysis.

**County-level survey tasks transmission subsystem**

A data center system will create survey tasks, including larger, high-resolution image data, but it is hard to transmit this information wirelessly to the survey fields. Therefore, in order to ensure agricultural statistical survey tasks being smoothly transmitted between the center system and field survey vehicle or terminals, the survey tasks will be transmitted to the county transmission subsystem at the county level via cable, LAN mode and then bridge to an agricultural statistical field survey vehicle. The transmission of survey data and results will follow a reverse order, bridging through the county-level survey tasks transmission subsystem.
Software and IT infrastructure (cloud computing platform)

A computer and network subsystem is the basic support of a spatial agricultural statistical survey. Because the geospatial agricultural statistical investigation system is centrally deployed, all computing devices and their corresponding software are deployed in the survey center, with each region of the operational application system node having its own independent running environment, although access to these resources would be via a client program. The centrally deployed environment mainly includes: computing devices and storage devices, and middleware and system software. Among them, the computing equipment is given priority to the server, the server computer system such as node deployment in information sharing and service platform. Each node statistics business with Client/Server (C/S) architecture design of computer system structure, the server is divided into two classes of data server, application server and portal server.

In order to configure, manage and maintain the servers, storage, network resources, and to secure remote sensing applications related to the efficiency and high availability of data processing, it is necessary to have dynamic management, virtualization, and ensure a high availability system of cloud computing platform. A cloud computing platform can be used to develop high-performance applications through the distributed computation, and distributed storage and system scalability, thereby meeting the needs of remote sensing mass data storage, processing and analysis requirements.

Challenges and future developments

Challenges

The challenges come from China’s farming system complexity.

- There are many crops varieties and regional differences in China, and the farming system is a non-scale type of farming, and crop rotation and intercropping are widespread.

- The landscape is also fragmented in most regions, except the northeast, and in a few other areas, the size of crop plots are very small.

- The terrain in China is complex, with crops being grown in all types of landscapes, including plains, hills, and mountainous areas.

- Market impact is important, with farming activities largely influenced by the market and rapid changes in cropping structure changes.
Land-use changes rapidly because of commercial activities.

**Other challenges include:**

- Time phase requirement of satellite imagery.
- Because crops are strongly seasonal, the acquisition of suitable remote sensing images in a specific period is not always possible, which highly limits the use of remote sensing area measurement surveys.
- Cloudy and rainy weather. Because of the constraint of cloudy or rainy weather and the timing of the satellite passing overhead, data availability is very limited for large-scale crop area remote sensing.
- Identification accuracy. The accuracy of remote sensing identification must be researched so that it improves substantially for use in agricultural statistics.

**Future development**

Along with continuous integration, innovation and improvement of survey technologies, the survey methodology and data base were consolidated. The efficiency of survey conducting and data quality was improved, and data service ability will be improved gradually. Finally, comprehensive remote sensing agricultural statistics will become operational.
ETHIOPIA

By Aberash Tariku Abaye

Introduction

Agriculture is the foundation of Ethiopia’s economy. According to the 2012/2013 Growth and Transformation Plan, agriculture contributes nearly 43% of Ethiopia’s gross domestic product (GDP). The Plan also reported that the country’s GDP grew by 9.7% and agriculture grew by 7.1% during these years. The main driver of the rapid economic growth has been the agriculture sector, in particular crop production, which accounted for about 30% of the GDP.

The Central Statistical Agency (CSA) has been conducting annual agricultural surveys for more than three decades through its integrated household survey programme. The annual agricultural survey conducted by CSA usually collects agricultural data on cultivated area and production for cereals, pulses, oilseeds, vegetables, root crops, fruits and other crops. Moreover, data on land and crop utilization, and agricultural practices were generated through the annual agricultural survey. The first agricultural census was conducted in 2001, and the second is planned for 2018. The agricultural survey conducted by CSA applies a list frame approach.

Master frame development is crucial for statistical activities. Any error in the master frame will result in inaccurate results. In Ethiopia, different approaches such as the list frame, area frame and community frame have been used to develop a master sampling frame.

Building and using master sampling frames: List frame approach

Methodology

The CSA has developed a master frame from population and housing censuses. A questionnaire was designed to collect data to be used for developing a master frame and the data was collected at the beginning of population and housing censuses. The questionnaire was filled out at the household level and summarized at the enumeration area (EA) level, and during both the cartographic and pre-enumeration phases. The major variables collected in the questionnaire
and then summarized are the number of: 1) agricultural households, 2) non-agricultural households, 3) total households, 4) small- and large-scale manufacturing enterprises, 5) cottage industries, 6) wholesale trade enterprises, and 7) retail and service trades. Data were collected at the household level by asking at each household whether there is a household member who is involved in these activities, and to check if there is an operator within the household. Scanning is used to capture the data in the listing form. The frame, which was made of EAs with the total number of households, was compiled directly from the census questionnaire, and the CSA now uses that frame.

For large-scale farms, a commercial frame was compiled. Farms classified as commercial include state farms, private commercial farms, cooperative farms and enterprise farms. The commercial frame is updated every year, and updating it involves using the list of farms from the previous year, which is sent to branch offices to determine whether they are still operational. Branch offices verify whether the commercial farms on the list are still operational, or not, by verifying on the ground the existence and operation of the commercial farms. Branch offices also cross check the list of farms with the investment offices to determine if there are any new commercial farms.

The master frame developed from population and housing censuses is used to select samples for the annual agricultural survey. In the list frame approach, stratified random sampling with zones as strata, EAs as primary sampling units (PSUs), and households as secondary sampling units (SSUs), is used. PSUs, which are EAs, are selected by probability proportional to size, with size being the number of households from population and housing censuses. Because we were not able to get a full and clean frame from the listing questionnaire administered at the beginning of the census, which includes the number of agricultural households, we decided to use the frame directly compiled from the census questionnaire, with total number of households. Households are then listed at the beginning of the survey, and sample households are selected. Once the sample households are selected, then data on cultivated area and production by crop type, land and crop utilization and agricultural practices, are collected. The same sample households are used to collect data during the long rainy season, short rainy season, forecast and livestock.

The CSA is planning to conduct an agricultural census in 2018. The first agricultural census was conducted in 2001, and was a large sample census that provided data at the lower district level. The master frame developed from the previous population and housing census was used as a frame for the first agricultural census. Then, the agricultural census was used as a master sampling frame for the annual agricultural survey until the recent population and housing census.
census was conducted. This has been updated by new population and housing census data.

The list frame facilitates the integration between different household surveys, is suitable for collecting socioeconomic data, and helps with diversified representation. This method does not require imagery. The list frame system is well established and has been used for many years. The difficulty with a list frame is that it is time consuming because households are distributed in the EA. Some holdings can be missed because their holding is dispersed. This makes supervision difficult. The difficulty with updating the master frame — due to frequent administrative boundary changes — is also a challenge.

**Estimation procedure**

The following formulas were used to estimate total area of land under a specific crop, and the production and yield of a specific crop in a stratum.

**For estimating total area of land under a specific crop:**

\[
A^*_h = \sum_{i=1}^{n_h} W_{hi} \sum_{j=1}^{H_{hi}} a_{hij} = \sum_{i=1}^{n_h} W_{hi} a_{hi}
\]

in which, \(W_{hi} = \frac{M_h H_{hi}}{n_h m_{hi} h_{hi}}\) is the basic weight.

Where:

\(h\) represents the stratum
\(n_h\) is the total number of sample EAs successfully covered in the \(h^{th}\) stratum.
\(M_h\) is the measure of size of the \(h^{th}\) stratum as obtained from the sampling frame.
\(m_{hi}\) is the measure of size of the \(i^{th}\) sample EA in the \(h^{th}\) stratum obtained from the sampling frame.
\(H_{hi}\) is the total number of agricultural households of the \(i^{th}\) sample EA in the \(h^{th}\) stratum.
\(h_{hi}\) is the number of sample agricultural households successfully covered in the \(i^{th}\) sample EA in the \(h^{th}\) stratum.
\(a_{hij}\) is the value of area for agricultural households \(j\), in the \(i^{th}\) EA in the \(h^{th}\) stratum under a specific crop.

\(a_{hi}\) is the sample total area under specific crop for EA \(i\) in stratum \(h\).

\(\hat{A}_h\) Estimate of total area under specific crop in stratum \(h\).

**For estimating total production under a specific crop:**

\[
\hat{P}_h = \sum_{i=1}^{n_h} w_{hi} P_{hi}
\]

in which, \(P_{hi} = a_{hi} \times \bar{Y}_{hi}\)

Where, \(\bar{Y}_{hi} = \frac{Y_{hi}}{16C_{hi}}\) is average yield per square meter of a specific crop in the \(i^{th}\) EA in the \(h^{th}\) stratum.

\(\hat{P}_h\) is the estimate of total quantity of production of a specific crop in the \(h^{th}\) stratum.

\(Y_{hi}\) is the sample total quantity of production of a specific crop from defined area of land for crop cutting of a crop in the \(i^{th}\) EA in the \(h^{th}\) stratum.

\(P_{hi}\) is the estimate of total quantity of production under specific crop for EA \(i\) in stratum \(h\).

\(C_{hi}\) is the number of crop cutting of a specific crop in the \(i^{th}\) EA in the \(h^{th}\) stratum.

**For estimating yield of a specific crop in stratum \(h\):**

\[
\hat{Y}_h = \frac{\hat{P}_h}{\hat{A}_h}
\]
Sampling variance of estimates:

Sampling variance for the estimate of stratum total of area, production and yield for a specific crop are estimated by the following formulas.

\[
\text{Var}(\hat{A}_h) = (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i=1}^{m} \left( \hat{A}_{hi} - \frac{A_h}{n_h} \right)^2 + f_h \sum_{i=1}^{m} (1 - f_{hi}) \left( \frac{h_{hi}}{h_{hi} - 1} \sum_{j=1}^{n} \left( \hat{A}_{hij} - \frac{A_{hij}}{h_{hi}} \right) \right)^2
\]

\[
\text{Var}(\hat{P}_h) = (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i=1}^{m} \left( \hat{P}_{hi} - \frac{P_h}{n_h} \right)^2 + f_h \sum_{i=1}^{m} (1 - f_{hi}) \left( \frac{h_{hi}}{h_{hi} - 1} \sum_{j=1}^{n} \left( \hat{P}_{hij} - \frac{P_{hij}}{h_{hi}} \right) \right)^2
\]

\[
\text{Var}(\hat{Y}_h) = \frac{1}{A_h^2} \left[ \text{Var}(\hat{P}_h) + \hat{Y}_h^2 \text{Var}(\hat{A}_h) - 2\hat{Y}_h \text{Cov}(\hat{P}_h, \hat{A}_h) \right]
\]

Where,

\[
\text{Cov}(\hat{P}_i, \hat{A}_i) = (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i=1}^{m} \hat{A}_i \left( \frac{\hat{A}_i}{n_h} \right) \hat{P}_i - \left( \frac{\hat{P}_i}{n_h} \right)
\]

\[
f_h = \text{average first stage probability of selection of EAs within stratum } h.
\]

\[
f_{hi} = \frac{h_{hi}}{H_{hi}} = \text{average second stage probability of selection within the } i^{th} \text{ sample EA in stratum } h.
\]

\[
\hat{A}_{hi}, \hat{P}_{hi} \text{ are weighted total area and production, respectively, of a specific crop in the } i^{th} \text{ EA and } h^{th} \text{ stratum.}
\]

\[
\hat{A}_{hij}, \hat{P}_{hij} \text{ are weighted values of area and production, respectively, from } j^{th} \text{ agricultural household in the } i^{th} \text{ EA and } h^{th} \text{ stratum under a specific crop.}
\]

Because all strata are independent, the total variance at the regional and country levels is computed by aggregating the result obtained at the zone/special wereda level:

\[
\text{Var}(\hat{A}) = \sum_h \text{Var}(\hat{A}_h), \text{Var}(\hat{P}) = \sum_h \text{Var}(\hat{P}_h) \text{and} \text{Var}(\hat{Y}) = \sum_h (\hat{Y}_h)
\]
Where, \( L \) is the number of strata (zone/special wereda).

In estimating the sampling variance by the above formula, selection of EAs within a stratum is assumed to be with replacement. By so doing, the variance estimate may be slightly over-estimated but it greatly simplifies the estimation procedure.

**Coefficient of variation of estimates:**

Coefficient of variation (CV) in the percentage of estimate of stratum total for area, production and yield for a specific crop is given by:

\[
CV(\hat{A}_h) = \frac{\sqrt{Var(\hat{A}_h)}}{\hat{A}_h} \times 100, CV(\hat{P}_h) = \frac{\sqrt{Var(\hat{P}_h)}}{\hat{P}_h} \times 100, CV(\hat{Y}_h) = \frac{\sqrt{Var(\hat{Y}_h)}}{\hat{Y}_h} \times 100
\]

**95% confidence interval (CI) of stratum total of area:**

\[
\hat{A}_h \pm 1.96 \times SE(\hat{A}_h)
\]

Where, \( SE(\hat{A}_h) = \sqrt{Var(\hat{A}_h)} \) is the standard error of the estimate of the stratum total of area.

Estimates of standard error and CI for the other estimates can also be calculated by adopting the above formulas.

**Building and using master sampling frames: Area frame approach**

**Methodology**

In the area frame approach, EAs are used as PSUs, and 40-hectare (ha) segments are used as SSUs. The CSA used two inputs to develop an area frame: enumeration area maps, and land cover maps.

The EAs (which are the PSUs) were delineated for the purpose of the population and housing census. The criteria to delineate an EA were that 150–200 households were within a rural area. A topo sheet was used as a base map and global positioning system (GPS) readings for the EA corners (turning points)
were plotted onto the topo sheet and then the EA map was then traced from the
topographical map. EAs were georeferenced.

The land cover classification activity is designed to produce a land cover
database that will provide a standardized, multipurpose product that is useful for
environmental and agricultural purposes. Satellite imagery, appropriate
software, and predefined legends are required for the land-cover classification.
The CSA used spot-5 satellite imagery, ARCGIS, Definiens\textsuperscript{13} and MadCat
software for land-cover classification. An appropriate legend was derived from
the standard legend prepared by the Food and Agriculture Organization (FAO).
The land is stratified into land-cover categories in the land cover classification
system, which is used as a basis for developing the area frame. The land cover
classification system is used to stratify the PSUs (or, EAs) based on their percent
of cultivated land.

To develop an area frame, the CSA-digitzed EA maps that were obtained from
census cartographic work, are overlaid on the land-cover map.

Four strata were created for the area frame based on crop intensity:

- Stratum I crop intensity 75% or more
- Stratum II crop intensity 50–74%
- Stratum III crop intensity 25–49%
- Stratum IV crop intensity less than 25%. This stratum was subdivided
  into two substrata in the recent area frame practice.

PSAs were selected by probability proportional to size, with size being the
number of segments. The sampled EAs were divided to 40-ha segments, with
two segments selected systematically from each EA for data collection. A closed
segment approach was used for data collection, whereby all of the fields (land
use) within the selected segment are listed, and the questionnaire is filled in for
each field. Commercial farms were treated separately and an independent survey
was conducted for them. Hence, a multiple frame approach, which is a
combination of an area frame and list frame, was implemented in the pilot
survey.

\textsuperscript{13} Definiens provides a simple-to-use iterative workflow for the ArcGIS environment to enrich
GIS applications with additional intelligence from images, thereby producing accurate and
detailed map data in a standardized and cost-efficient way.
The second step is to estimate the percent of crop land (crop intensity) for each EA, and this is calculated using the interpreted imagery and appropriate software.
Stratification is done by forming groups of similar EAs with respect to crop intensity. The four strata created based on crop intensity as indicated above.

### ENUMERATION AREAS CLASSIFIED INTO STRATA BASED ON THE PERCENT OF CROPLAND FROM LAND COVER.

- **0 TO 24**
- **25 TO 49**
- **50 TO 74**
- **75 AND ABOVE**

Allocated numbers of EAs from each stratum are selected by probability proportional to size, and sampled EAs are divided into 40-ha segments. These segments are treated as SSUs.
An enumeration area that has been subdivided into seven, 40-ha secondary sampling units.

The formation of the 40-ha segments is based on clear identifiable physical boundaries as much as possible. Two segments are selected systematically from each EA for data collection.

Segment number 2 selected for the enumeration area (or primary sampling unit) shown above.
Estimation procedure

Estimate of total in each stratum is given by:

\[ \hat{Y}_h = \sum_{i=1}^{n_{ea}} \sum_{j=1}^{n_{hi}} W_{hi} Y_{hij} \]

The weight for each segment is calculated as:

\[ W_{hi} = \frac{M_{sh} N_{hi}}{n_{ea} M_{ea} n_{hi}} = \frac{M_{sh}}{n_{ea} n_{hi}} \]

The variance for the estimates in each stratum is calculated as:

\[ Var(\hat{Y}_h) = \frac{n_{ea}}{n_{ea} - 1} \sum_{i=1}^{n_{ea}} \left( \hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \]

\[ V(\hat{Y}) = \sum_{h=1}^{i} var(\hat{Y}_h) \]

Area frame survey saves time because holdings are close to each other. It also avoids missing fields. The data can also be cross checked with the total area of the segment. An area frame approach also facilitates supervision. On the other hand, identifying the owner of the field may take some time. In addition, an area frame requires satellite imagery and land cover classification, which requires a large budget. All aspects of the method should be examined carefully because the area frame system is not yet well established. In the area frame design, selecting the appropriate approach for the area, production and other socioeconomic information should also be well planned out.
Results obtained and precision of estimates

Estimates of the area and CV for the Oromia region for major crops from the 2011 area frame data are shown below.

<table>
<thead>
<tr>
<th>Region</th>
<th>Crop</th>
<th>Area</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>OROMIA Region</td>
<td>Barley</td>
<td>472 631</td>
<td>11.00</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>1 248 631</td>
<td>11.00</td>
</tr>
<tr>
<td></td>
<td>Sorghum</td>
<td>1 118 824</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Teff</td>
<td>1 517 239</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Wheat</td>
<td>859 627</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Estimates and CVs at the regional level include Borena and Guji zones.

This table shows the estimates of area and CV at the regional level. Because the sample size is calculated to obtain a reliable estimate at the regional level, it will be more practical to compare estimates at the regional level. Borena and Guji zones are those zones where the majority of residents are nomads, and the weights in these zones are highly abnormal.

Lessons learned

The following lessons were learned from implementing the area frame:

- Clear satellite imagery is needed.
- Land cover classification, based on satellite imagery, is required. This needs appropriate software and skill for interpretation.
- GPS for delineating and measuring fields is required.
- A segment map with clear boundaries needs to be prepared.
- Preparing a clear manual, and handling intensive training is important
- Enough time should be allocated for data collection.
- When the segment boundary dissects the field, care should be taken to measure only part of the field within the segment.
- Commercial farms should be identified and treated separately.
- Care should be taken to avoid merging adjacent fields during data collection.
- The impact of abnormal weights in some strata needs to be determined.
- The treatment of strata with less cropland should be dealt with.

Collecting socioeconomic data, including data on livestock and integration with other socioeconomic surveys that use a list frame, is one of the major issues to be considered in the application of an area frame.
Comparison of results from an area frame and list frame.

<table>
<thead>
<tr>
<th>Region</th>
<th>Crop</th>
<th>Area frame</th>
<th>List frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Area estimate (ha)</td>
<td>CV</td>
</tr>
<tr>
<td>Oromia Region</td>
<td>Barley</td>
<td>472 631</td>
<td>11.00</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>1 248 631</td>
<td>11.00</td>
</tr>
<tr>
<td></td>
<td>Sorghum</td>
<td>1 118 824</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Teff</td>
<td>1 517 239</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Wheat</td>
<td>859 627</td>
<td>8.00</td>
</tr>
</tbody>
</table>

A comparison of estimates at the regional level for an area frame and list frame was done. Estimates of major crops (except sorghum, coffee and chat) compare reasonably well between area frame and list frame estimates. Stratum 1 and Stratum 2 work well, but Stratum 3 and Stratum 4 have high weights and contribute towards higher variability. In addition, stratification needs to be modified in view of special problem crops such as coffee. In particular, Stratum 4, which is very large, needs further substratification. A comparison of CVs for area frame and list frame was done at the regional level. It is acknowledged that CVs are higher in an area frame, with Stratum 4 contributing more towards CVs. Some 239 EAs were selected, with two segments per EA, with this sample size estimation, were targeted. CVs at the regional level are around 10% for an area frame. For zone-level reliable estimates, sample sizes will be much higher; or else, one has to be prepared for zone-level estimates with very high CV levels.

**Development of a community frame**

The community questionnaire was designed to collect data on agricultural- and health-related issues such as agroecology, types of crops grown in *meher* (long rainy season) and *belg* (short rainy season), culture of growing rare crops, irrigation and types of major diseases in the EA. Data were collected only from rural areas during population and housing censuses. Respondents in the community questionnaire were representatives in the EA, from three to five people, and usually elders, officials and development agents. For the community questionnaire, data were captured by data entry instead of scanning. Data entry was done after editing, which minimizes mismatch.

The community questionnaire is useful for stratification purposes for surveys. The information collected in community questionnaire is based on respondents’ knowledge, which may create discrepancy with the data collected on a scientific basis.
Survey data collection

Agricultural data from the annual agricultural survey are collected from the entire country except from non-sedentary zones of two regions. In total, 2142 EAs were selected in 2012/2013 for the survey. In the list frame approach, measuring tapes and a compass were used for area measurements and GPS was used in Oromia Region. Field staff was estimated to be around 2 200 enumerators and 572 supervisors. The ratio of supervisors to enumerators was 1:3. In total, ETB (Ethiopian birr) 90 million, equivalent to USD 4.5 million, has been allocated for the annual agricultural survey each year.

In the area frame approach, the enumerator at the beginning identifies the segment boundary using a segment map and GPS coordinates. GPS is used to delineate a segment and to measure all of the fields within the selected segment. The sample size for the Oromiya region area frame was 215 PSUs (EAs) and 430 SSUs (two segments per EA). This sample size was determined based on the precision and sample size in the West Shoa area frame pilot. These 215 EAs were allocated to the 17 zones proportional to the cultivated area. A two-stage training was conducted. The first stage included training for trainers, and then enumerators and supervisors were then trained on data collection. The training included both classroom training and field practices.

Future plans

The CSA will conduct the fourth population and housing census in 2018. Cartographic work for the census began in 2015. GPS-enabled personal digital assistant will be used to do the cartographic work. This method will enable the CSA to have an electronic copy of the EA maps automatically, and to obtain GPS readings for each household. This method is expected to build a better master sampling frame with GPS readings for each household.
References


GEORGIA

By Maia Guntsadze and Vasil Tsakadze

Summary

This paper reviews the experience of Georgia in building and using the master sampling frame for producing agricultural and rural statistics. In November 2014, the National Statistics Office of Georgia (Geostat) conducted an agricultural census together with a general population census. The data derived from the census has been used as a master sampling frame.

The census was conducted using global information system (GIS) maps; therefore, all of the data derived from the census are linked to the GIS data. Geostat also conducted an agricultural census for legal entities in order to have full coverage of agricultural census data (including common land data).

This paper describes the design process of the master sampling frame, and the agricultural survey that was conducted based on this frame. The master sampling frame obtained from the census is very comprehensive and can be used for many different surveys because it was derived from the integration of population and agricultural censuses. The benefits of conducting agricultural and population censuses together, in building the master sampling frame, are described. Also described is a core set of variables that were used while building the master sampling frame and the main indicators derived from the survey.

Introduction

The statistical system in Georgia is centralized. The statistical law of Georgia creates a general framework for the country’s statistical system and sets coordination principles for all institutions responsible for producing official statistics in Georgia. According to the law, the production and dissemination of statistics is based on the 10 basic United Nations principles of official statistics of the European Statistics Code of Practice and internationally recognized basic principles of statistics. Geostat is a coordinating body of Georgia's statistical system and the only producer of official agricultural statistics in the country.14

14 The Georgian statistical system covers the whole area of the country, except the occupied territories.
Geostat is also the agency responsible for conducting population and housing censuses, and agricultural and economic censuses.

Georgia has substantial agricultural potential. The government places great emphasis on the need for investing in increasing output and productivity in the agricultural sector. The demand for agricultural data is increasing daily, and Geostat must respond to users’ needs.

Agriculture has always been one of the important sectors of Georgia’s economy. Despite the fact that the share of agriculture, hunting, forestry and fishing in terms of gross domestic product (GDP) was only 9.2% in 2015, 48.6% of employees in Georgia were employed in this sector. The proportion of the population living in rural areas is 42.8%, according to the 2014 population census.

Almost all households living in villages are agricultural holdings, and even in small towns two-thirds of households are engaged in agricultural activities. Overall, three out of five households are engaged in agriculture and these households are scattered throughout every region of the country. According to the Census of Agriculture 2014, the total number of agricultural holdings is around 642,000; out of this, only 2000 are legal entities while the other 640,000 are agricultural households. The majority of agricultural holdings are small and produce agricultural products for their own consumption.

Due to its favorable geographical location and climate, Georgia produces more than 25 kinds of permanent crops and more than 20 kinds of annual crops. In addition, animal husbandry is quite common. One of the characteristics of the agricultural sector is that it is not well-specialized and the majority of holdings produce many different kinds of agricultural products.

The main source of current agricultural statistics is the Sample Survey of Agricultural Holdings, which dates back to 2007. Its methodology was developed with the support of the Food and Agriculture Organization and the United States Department of Agriculture. A sample frame of the survey was based on the 2004 agricultural census, which was the first in the history of independent Georgia. In 2014, Geostat conducted the second agricultural census in combination with the population census. Based on the census database, in 2015–2016, Geostat prepared a master sampling frame for agricultural surveys.
Methodology for building the sampling frame and designing the survey

Type of frame chosen

The type of master sampling frame is the list frame, due to the fact that the main sources of information used to create a master sampling frame are population and agricultural censuses. Because the general population and agricultural censuses covered only the household sector, an Agricultural Census for Legal Entities was conducted separately. To ensure full coverage of the list of potential agricultural enterprises, all existing reliable sources in the country were used, including:

- existing sample frame of the regular quarterly survey on agriculture;
- statistical business register;
- public register;
- information available from the Ministry of Agriculture;
- information available from local governments;
- information available from the Ministry of Education; and
- information available from the Patriarchate of Georgia.

As a result, the census database contains 642,209 agricultural holdings, of which 2,246 are agricultural legal entities, and 639,963 are family holdings.

Within the census database, each household has been assigned a unique identification number, as has each legal entity. In the database, the following variables were used in order to form the master sampling frame:

- unique identification number of holding;
- status of holding (i.e. legal entity or household);
- holding address and residence or office address;
- areas of various land categories such as: arable land, temporary crops sown, uncultivated land, permanent crops, orchards, land under berries, vineyards, citrus plantations, tea plantations, nurseries, other permanent crops, greenhouses; natural pastures and meadows, non-agricultural land, reservoirs for aquaculture.
number of cows, buffaloes, pigs, sheep, goats, horses, mules and donkeys, rabbits, poultry, and colonies of bees;

- number and use of agricultural machinery; and

- variables on demography, housing conditions, emigrants, aquaculture holdings, greenhouse holdings, municipal (i.e. common) land.

**Detailed description of building the frame**

In November 2014, the National Statistics Office of Georgia conducted the census of agriculture together with the population census. In fact, four different censuses were conducted together: housing census, population census, census of emigrants, and agricultural census. Additionally, the census identified holdings operating aquaculture and greenhouses. The census was conducted using GIS maps.

Geostat started preparatory work for the general population census in 2006. One of the principal parts of this preparatory work consisted of implementing GIS in order to create digital cartographic material and use them in the data collection and dissemination process. The GIS unit of Geostat developed GIS maps using cartographic materials available in the country. A supplementary work was done to cover the territories where cartographic material was missing. In addition, a numbering script was prepared in order to give a unique code to each parcel and buildings located on it. After this, supplementary work included updating existing software and switching to a fully automated system.

In 2013, Geostat conducted preliminary field work to elaborate the list of dwellings and households existing in Georgia. In order to digitize the collected information and to integrate it with the cartographic data, a dedicated software was developed, which used the coding system created by the GIS unit of Geostat. The information received from the preliminary field work was used for correcting and finalizing existing data on buildings, and for effectively distributing work among census enumerators. In addition to these, household duplication and “missing” risks were minimized. By implementing GIS, Geostat managed to significantly decrease the time needed for preparing cartographic materials, use financial means in a more effective way, and prepare high-quality maps.
Preparing the census questionnaires was undertaken in accordance with international guidelines and recommendations, taking into consideration local users’ needs. The census questionnaire consisted of four separate parts for the four different censuses: Questionnaire on Dwelling (Q1), Personal Questionnaire (Q2), Questionnaire on Emigrants (Q3), and Questionnaire on Agriculture (Q4). Each questionnaire had a unique identification number that linked them to each other. In total, the census questionnaire was nine-pages long: two pages for Q1, two for Q2, one for Q3, and four for Q4. Q1 contained questions on dwelling conditions and the full list of household members living in the dwelling, including their status (e.g. usual resident, temporary absent, temporary present, emigrant). The questionnaire used (Q2 or Q3) depended on the household member’s status.

Because the population and agricultural censuses were conducted together, it was agreed to use the same reporting unit; thus, a household was considered as one agricultural holding. In the case of two or more holders in the same household, one of them was selected as a holder and their operated lands and livestock were summed up and assigned to the holder. In the case of one or more households in a holding, each household was considered as a separate holding, and land and livestock were assigned proportionally according to the share of households in the holding.

There were main identification questions in the Q1 for agricultural holdings. According to these questions, if a household was identified as a holding, Q4 was filled out by the enumerators. Besides, Q1 identified the holder. Additionally, Q1 contained identification questions on holdings engaged in aquaculture activities. The questionnaire on agriculture (Q4) did not contain any questions about the holder other than the holder's unique number (as indicated in Q1 and Q2) because all relevant information was covered by Q1 and Q2. Q4 covered all core indicators recommended by FAO, and local users’ needs were taken into account.\(^{15}\)

For the Agricultural Census for Legal Entities, an electronic questionnaire was prepared, which was also used for the census of municipalities (common land data). The content of the electronic questionnaire was the same as for Q4. In addition to the questions similar to Q4, there were additional questions regarding personal characteristics of the company director or holder.

---

\(^{15}\) [http://census.ge/en/methodology/kitkhvarebi]
The end result is a census database that contains information on dwelling conditions, personal information on the population, personal information on emigrants, agricultural data on holdings, personal information on holders, data on households operating aquaculture and greenhouses, and geographic data. The database is a comprehensive source for analysts and researchers to cross-analyze socioeconomic and agricultural data. The coding system included in the census database easily links all data and makes it easier to create a master sampling frame for agriculture.

**Details of the sampling design**

Holdings were partitioned into strata using the following information:

- region,
- size class of holding (calculated according to agricultural index of holding), and
- type of holding (e.g. family holding or legal entity).

Regional partitioning and enumeration was done by using the same codes that were used in the agricultural census. Holdings were grouped by the following size classes: 1 = small, 2 = medium, 3 = large, 4 = very large. In each region, all small holdings were grouped into one stratum. The same was done for medium size holdings and for very large holdings.

As for the large holdings, in each region, two strata were created: one for large households and one for large legal entities. This was essential because it was found that large legal entities are drastically different in their structure from large households. In particular, the average agricultural index for large legal entities is 40.6, while for households it is 14.4. So, holdings in each region were grouped into five strata: small, medium, large legal entities, large households, and very large holdings. In total, all holdings in Georgia were partitioned into 55 strata.

Holdings in each stratum were grouped into clusters. A cluster is a group of holdings situated in one territorial area. The clustering process consisted of several stages. In the first stage, the main parameter of clustering was settlement. Holdings from the same populated area, which belonged to the same stratum, were grouped into one cluster. In the second stage, huge clusters were partitioned into smaller ones. That particular process used the same partitioning design as in the population census, in that huge clusters were partitioned into enumeration areas. At the last stage, very small clusters were unified with neighbor
In total, the sampling frame was partitioned into 11311 clusters.

Currently, the sample includes approximately 12,000 holdings, which are allocated into strata. There are 420 holdings within the “very large” stratum, and all are included in the survey. In order to obtain robust estimations for each region, from the rest of the sample, 9,000 units were allocated among regions according to the Neiman allocation principle, which gives the best estimations for the total population. The remaining 2,580 units were allocated among small regions so that in each region there is the minimal sample size needed for robust estimations. The assigned quantities for each region were allocated among strata according to the Neyman principle.

In the survey, two-stage cluster sampling was used. Sample groups were formed independently in each stratum. In the first stage, within each stratum, clusters were selected. In the second stage, holdings were selected from each cluster’s holding list.

Very large holdings are included in the survey with a probability of 1; in other words, all clusters of very large holdings and all holdings within those clusters are included in the survey.

In strata other than very large holdings, the first-stage sampling unit is the cluster. In each sampled cluster, the same number of holdings is interviewed. In particular, in sampled clusters from small size holdings, twelve holdings are interviewed, in sampled clusters from medium size holdings strata, eight holdings are interviewed, and in sampled clusters from large size holdings, four holdings are interviewed.

In each stratum the number of sampling clusters was calculated by dividing the number of sampling agricultural holdings by the number of holdings in each sampling cluster of this stratum. In each stratum clusters were sampled using probability proportional to size method.

As mentioned, the second-stage sampling unit is the holding. In the sampled clusters, the sampling of holdings was done using the random sampling method. Afterwards, because of rotation, by the time the new sample of holdings is needed, those holdings that occupy the next place in the list will be chosen.

One important feature is the rotational plan. The panel survey design is used and sampled holdings will remain in the survey for three years. After that, they will be changed by other holdings of the same strata. All 420 very large holdings will
continuously be included in the survey. The other 11,628 sampled holdings are grouped into three rotational groups. All sampled holdings of the cluster belong to the same rotational group. Each rotational group covers approximately the same number of sampled holdings. In particular, in the first rotational group there are 3,885 holdings, in the second rotational group there are 3,879 holdings, and in the third group there are 3,864 holdings.

Firstly sampled holdings will not remain in the survey for three years. In 2017, the first rotational group will change; in 2018, the second rotational group will change, and the third rotational group will change in 2019. Exiting holdings will be substituted by the holdings next to them on the list of cluster holdings. When there are not enough holdings on the list, the cluster is considered to be “expired” and is substituted by a randomly selected cluster of the same stratum. The newly entered holding belongs to the same rotational group as the preceding one.

**Formulas for estimators**

**Calculation of estimators**

In order to obtain aggregated estimators from the survey, a weighting process is used. Weighting is conducted at the stratum level. Each holding within the stratum has the same weight. At first, the initial weight of each sampled holding from s-th stratum is calculated by the formula:

\[
W_s^{(0)} = \frac{N_s}{n_s}
\]

Where, \(N_s\) is the number of holdings and \(n_s\) is the number of sampled holdings in the s-th stratum. Each enquired holding from s-th stratum has a weight:

\[
W_s^{(1)} = \frac{N_s - u_s \cdot W_s^{(0)}}{r_s}
\]

Where, \(r_s\) is the number of answers in this stratum, and \(u_s\) is the number of sampled holdings that no longer exist.

In the very large holdings strata, the difference between the holdings sizes may be very high and the allocation of rejection weights among enquired holdings may result in a “bad” output. Thus, post-stratification is needed for such holdings. To begin, the main specialization of these holdings must be determined and this is done by identifying the activity (crop or livestock) that contribute the most to the holding’s agricultural index. All very large holdings with the same
specialization are grouped into the same stratum. One stratum is formed for particularly large holdings, whose agricultural index is more than 1 500 (in the sample, the number of such holdings is 19). Also, the holdings that are unique by their specialization are grouped in the same stratum. Enquired holdings from this stratum always have a weight equal to 1, \( W_s^{(0)} = 1 \), while existing holdings that have not been interviewed (for whatever reason), the data are created individually for each holding, based on the imputation method and the information received from the previous survey or from other sources. The weighting process for the rest of the very large holdings is conducted in the same way as with small, medium and large holdings.

**Results**

The annual electronic publication “Agriculture of Georgia” contains information on various aspects of agriculture within the country. The publication contains four sections. Section 1 includes general information on Georgian agriculture, such as rural population, share of agriculture in the GDP, and the structure of agricultural output and volume indices. Section 2 presents nationwide and regional data on areas under annual crops, harvested areas, crop production and yield. Section 3 contains variables on animal production and livestock numbers, while Section 4 presents food balance sheets (supply and utilization) for wheat, maize, potatoes, vegetables, grapes, total meat, beef, pork, sheep and goat meat, poultry meat, milk and milk products, and eggs. In particular, the following estimators are obtained from the survey:

- Sown area, harvested area, yield and production of annual crops (wheat, barley, maize, haricot beans, potatoes, tomatoes, cucumbers, red beets, cabbages, capsicums, paprika, garlic, onions, greens, carrots, eggplants, pumpkins, watermelons, melons, annual grasses, perennial grasses).

- Production of permanent crops (apples, pears, quinces, plums, cherries, apricots, peaches, sour plums, walnuts, hazelnuts, pomegranates, persimmons, mulberry, mushmalas, feijoa, kiwis, berries, grapes, tangerines, oranges, lemons, tea).

- Shares of households and legal entities in sown areas and in crops production.

- Number of livestock (cattle, including buffaloes, dairy cows and dairy buffalos, pigs, sows sheep, ewes, goats, mother goats).

- Number of poultry.
• Number of beehives.

• Livestock production (meat: beef, pork, sheep and goat meat, poultry meat; milk: cow’s milk, sheep and goat milk; eggs; wool and honey).

• Shares of households and legal entities in livestock numbers and in livestock production.

• Average yield of milking cows and buffaloes.

• Average clip per sheep.

• Litter of calves, buffalo calves, suckling pigs, goat kids and lambs.

• Average litter of calves and buffalo calves per 100 dairy cows and dairy buffaloes.

• Average litter of suckling pigs per 100 sows.

• Loss of cattle, pigs, sheep and goats and poultry of all kinds.

• Use mineral fertilizers and pesticides.

In the annex, there are two result tables that have been published on Geostat’s website. Final results for all estimators was published on 16 June 2017.16

Resources mobilized and other cost

Costs of building the sampling frame

The total cost of the censuses is GEL 14,150,900.17 Expenses on employees’ compensation are GEL 11 326,400, of which GEL 289,800 were allocated to administrative and financial group, GEL 245,100 to organizational and methodological group, GEL 1,735,100 to GIS group, GEL 1 774,000 to data processing and software, and GEL 7 282,300 to field work.

Business travel expenses amounts to GEL 240,600. Needed services and equipment took GEL 2 583,900. Of which, an advertising campaign – GEL 671,100, stationery – GEL 68,500, uniforms – GEL 200,400, printing work – GEL 276,200, communications – GEL 14,900, office equipment and inventory

16 These are available at: http://geostat.ge/index.php?action=page&p_id=245&lang=eng
17 Equivalent to about USD 8 423,200 using the exchange rate in November 2014.

### Costs of conducting survey

Total cost of conducting Sample Survey of Agricultural Holdings in 2016 was GEL 868,200. It was allocated among initial, quarter and final enquiries. Total budget were distributed among the following components: interviewers’ compensation GEL 664,400, coordinators’ compensation GEL 87,400, data entry, logical controls and editing process GEL 43,600. In addition, compensation was paid for 4 contracted staff the total amount of GEL 55,900.

### Human resources mobilized

The total number of employees involved in the general population and agricultural censuses for 2014 was 12,200 workers. During the preparatory work in 2013–2014, 3,800 employees were involved, of which 1,600 were involved in 2013 and 2,200 in 2014.

The agricultural survey work involved 17 office staff, more than 350 interviewers and more than 80 coordinators. It should be noted that some coordinators were also interviewers, so the number of employees in the agricultural survey without their status is approximately 370 workers.

### Survey data collection

The agricultural survey included questionnaires for family holdings and legal entities. With households, paper questionnaires were used, while online questionnaires were used for legal entities. The contents of both questionnaires were the same. Three types of questionnaires were used: 1) inception enquiry, 2) quarterly enquiry, and 3) final enquiry.

The process of selecting new interviewers was done according to Geostat’s rules used by heads of regional offices. An authorized person from the Agricultural and Environment Statistics Division checked coordinators’ and interviewers’ knowledge level regularly by conducting tests. Every year, field staff are trained and additional training is conducted as needed.

After the interviewers are ready, the Agricultural and Environment Statistics Division distributed survey tools and the list of sampled holdings to the regional offices, where they were in turn distributed among municipalities, which are
allocated among interviewers through the coordinators. Each interviewer receives the list of only those holdings that she or he must visit in order to conduct an interview.

A face-to-face interview was conducted by interviewers with respondents using the appropriate questionnaire: inception, quarterly or final. In the case of legal entities, the coordinator helped the contact person from the legal entity to fill out the online questionnaires. In rare cases, when the contact person could not fill out the online questionnaires, the coordinators\(^{18}\) were allowed to conduct a face-to-face interview or phone call interview with them using the paper questionnaires.

In general, the interviewer (coordinator) tried to conduct the interview with the holder. If this was not possible, then the coordinator tried to talk with the second competent person in the household in a legal entity.

**Quality control**

Every year, several regions are selected for monitoring. Interviewers are chosen from these regions and their list is delivered to the monitoring division (with the randomly selected holdings that were interviewed by them) in order to check how appropriately the above-mentioned interviewers conducted their interviews (i.e. to determine whether the interview was conducted by the interviewer within the holding). In addition, the special questionnaire is filled out by the monitoring division. The data are compared to the data provided by the interviewer, and sections of the questionnaire are analyzed as to whether there were systematic mistakes made by the field staff.

The staff of Agricultural and Environment Statistics Division checked the relevance of the collected data and “cleaned” the database. Computer programmes were used to identify illogical and suspicious data. Based on phone calls with the contact persons, who were respondents during the interview, or with field staff, corrections were made. In addition, automatic imputation was done.

\(^{18}\) If coordinators were unable to conduct interviews, then the region chairman, under approval of the head of the Agricultural and Environment Statistics Division, entrusted the interviewer to contact the person of the agricultural legal entity.
Issues and challenges encountered, and future developments

The main challenge for Geostat was to keep the master sampling frame up to date. All reliable sources in the country are used to permanently update the frame. It is still problematic to update small-scale holdings because there is no farm or livestock registers in the country. The number of holdings during two agricultural censuses (2004 and 2014) decreased drastically. The Sample Survey of Agricultural Holdings during the period 2007–2015 was based on the 2004 census. Reconciling the data during the censuses was quite important and was demanded by the users. Unfortunately, there is lack of international practice and methodologies on reconciliation. Geostat is working on this issue and considering ways of data reconciliation with international organizations and experts. The final decision on reconciliation will be done based on analyzing different methods and the results derived from using them.

One important issue, which was beyond Geostat’s ability, was the problem of country coverage because of the uncontrolled territories. It is well known that considerable areas of land in the country are considered as occupied territories, thus holdings from these areas were not included in either the censuses or the survey.

The existing master sampling frame was used for this survey and can be used for various other surveys, such as integrated agricultural surveys, specific agricultural surveys, aquaculture surveys, greenhouse surveys, and socioeconomic surveys.

In 2015, Geostat prepared the “Strategic Plan for Agricultural, Environmental and Rural Statistics in Georgia 2016–2020” (SPAERS), and was supported by FAO under the Global Strategy project. It determines the planned activities regarding improving agricultural, environmental and rural statistics, including new surveys to be conducted in the coming years. The master sampling frame will be the main tool used for conducting agricultural surveys and will be used as a frame for making samples.

A master sampling frame facilitates the identification and monitoring process in the field because it contains detailed personal data of the holder and their household members, as well as GPS coordinates of the household’s location. It also facilitates the field work of enumerators by indicating where sampled units are located.

The master sampling frame for Georgia covers various variables regarding family holdings, agricultural enterprises, households, emigrants, housing
conditions, aquaculture holdings, greenhouse holdings, municipal (common) lands, and others.

Conclusions

By conducting the population and housing census together with agricultural census, Georgia’s National Statistics office has created the first master sampling frame for agricultural surveys. The identification codes system embedded in the questionnaires easily linked the population and agricultural censuses together, and hence, linked agricultural holdings data to individual household data. Using GIS maps during the census provided the master sampling frame with detailed geographic data.

The master sampling frame is used for regular agricultural surveys and will be used for one-time surveys that are planned for the coming years. Merging different surveys using a master sampling frame decreases survey costs and respondent burdens, which is crucial for a statistical system. A master sampling frame also facilitates achieving the planned activities determined by SPAERS in a more efficient way, and is a good tool for surveys studying agricultural activities for specific social groups. A master sampling frame can be widely used for the identification and monitoring process during field work, which will decrease the time needed for data collection and increase data quality.

It is important that the master sampling frame be updated regularly using data obtained from all agricultural and socioeconomic surveys based on the master sampling frame and on the reliable administrative sources available in the country in order to keep the frame up to date.
References


### Annex 1

#### Table 1

<table>
<thead>
<tr>
<th>Livestock and poultry numbers (end of period, ths. heads)</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Q</td>
</tr>
<tr>
<td>Cattle</td>
<td>1,109.8</td>
</tr>
<tr>
<td>Cows</td>
<td>545.0</td>
</tr>
<tr>
<td>Pigs</td>
<td>172.1</td>
</tr>
<tr>
<td>Sheep and goats</td>
<td>1,232.2</td>
</tr>
<tr>
<td>Poultry</td>
<td>8,413.3</td>
</tr>
</tbody>
</table>

#### Table 2

<table>
<thead>
<tr>
<th>Production of main animal products during the period</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Q</td>
</tr>
<tr>
<td>Meat (slaughtered weight, ths. tons)</td>
<td>14.1</td>
</tr>
<tr>
<td>Milk (mln. litres)</td>
<td>97.3</td>
</tr>
<tr>
<td>Eggs (mln. units)</td>
<td>151.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Production of main animal products during the period</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Q</td>
</tr>
<tr>
<td>Meat (slaughtered weight, ths. tons)</td>
<td>8.5</td>
</tr>
<tr>
<td>Milk (mln. litres)</td>
<td>92.8</td>
</tr>
<tr>
<td>Eggs (mln. units)</td>
<td>146.9</td>
</tr>
</tbody>
</table>
GUATEMALA

By Marino Barrientos

Introduction and background

According to Guatemalan law, the National Institute of Statistics is the institution responsible for generating official statistics on the country. In the agriculture sector, four national censuses of agriculture were conducted in 1950, 1964, 1979 and 2003. Between census periods, a set of surveys based on several types of frames were done by sampling.

Between 1965 and 1973, the General Directorate of Statistics (DGE, Dirección General de Estadística, by its acronym and name in Spanish) conducted several surveys of sampling using a list frame based on the information from the first National Farming Census, which was conducted in 1950, and the cartography used in the second National Farming Census was done in 1964.

Between 1974 and 1978, the DGE used information generated by the second National Farming Census in 1964, and the cartography used for the population and habitation census of 1973 came from the farming surveys.

Other studies were done in 1984 and 1985 by the National Institute of Agricultural Marketing (INDECA, Instituto Nacional de Comercialización Agrícola), on the basis of the information generated by the third National Farming Census in 1979. The sample of 1984 was used during 1986, 1987 and 1988.

In 1995 and 1996, the Ministry of Agriculture, Livestock and Food (MAGA, Ministerio de Agricultura, Ganadería y Alimentación) conducted a national survey of grains and a survey of farming, based on the information of the 1979 census and the cartography used for the 1994 population census.

In 2003, the National Institute of Statistics, with technical support from FAO, conducted the fourth National Farming Census, and designed a system of continuous farming statistics. The process started with the building of a sampling frame using an area frame, using satellite images and aerial photograph with stratification. This frame was combined with the list of big farms that the census identified during the surveys of 2005, 2006, 2007 and 2008 and based on
multiple frames (area and list). The design of the sample used for these surveys was stratified with two selection stages: Primary Unit Sampling (UPM, Unidad Primaria de Muestreo) and sampling segments within the UPMs.

The 2013 survey was carried out by a coordinated effort between the Instituto Nacional de Estatistica (INE) and the Ministerio de Agricultura, Ganaderia y Alimentacion (MAGA) with technical assistance from FAO. This methodology is also used currently, and its use is intended to increase in order to obtain more reliable information for several relevant variables in more disaggregated geographical levels and for other variables. The methodology can also be used for evaluating the effects of droughts and floods, and for supporting rural development programmes in the field.

**Survey objectives**

**The primary objectives of the 2013 survey were to:**

- determine the proportion of sowed land of priority crops during the period May to December for the entire country;
- collect information on the performance of priority crops during the period May to December; and
- estimate the production of priority crops, at the national level, during the period May to December.
Process flowchart

Generation of the sampling frame national territory (areas frame) 108 889.15 km²

Stratum delimitation  →  Segments delimitation

Sample design (areas frame) 1 500 segments

Estimators generation  →  Estimators generation

Sample design (areas frame) 1 500 segments

Criteria definition of farms size

Crops and species definition

Directory generation

Estimators generation  →  Estimators generation  →  Estimators generation  →  Estimators generation

Preparation of cartographic materials, cards, forms, handbooks and measuring equipment (e.g. GPS)

Recruitment, selection and training for field people

Collect field information in segments and directory farms

Consistence of the field information

Segments database generation  →  Farms database generation

Results generation (estimators by interval)
Methodology used for building and using the sampling frame

The universe of reference was defined, according to the National Farming Survey, as every surface within the national territory that is used for farming production or has the potential for production. This was done by classifying the national territory into two land types: agricultural and non-agricultural.

The non-agricultural category included populous centers as well as: educational, recreational and military facilities; jails; industrial estates; airports; shores; cemeteries; water sources; wetlands with forest and other vegetation; swamps; arid and mining areas; beaches; volcanic cones; national parks; and protected areas.

<table>
<thead>
<tr>
<th>Table 1. Guatemalan land area by category.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>National territory</td>
</tr>
<tr>
<td>Non-agricultural use</td>
</tr>
<tr>
<td>Agricultural use</td>
</tr>
</tbody>
</table>
Process for building the area sampling frame

1. Using Arc-Gis 9.3, all of the national territory was divided in squares of 100 ha, and 110 128 segments were generated.

2. Based on land cover and a land-use map in 2010, the percentage of the area was determined within each segment that had agricultural (including pastures) and was stratified according to the percentage of the agricultural area within each square. Four squares were generated.

---

19 1 hectare = 10 000 m²
3.

4. After the stratification, and depending on the size of the crop fields, a digital division of the squares was done, giving as a result the following segments’ sizes:

Table 2. Definition of strataums according to the cultivated surface of the segments

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Definition criteria of the strataums</th>
<th>Segment size (hectare)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Cultivated surface greater than 60% and large fields</td>
<td>25.00</td>
</tr>
<tr>
<td>B</td>
<td>Cultivated surface greater than 60% and small fields</td>
<td>6.25</td>
</tr>
<tr>
<td>C</td>
<td>Cultivated surface between 20% and 60%</td>
<td>50.00</td>
</tr>
<tr>
<td>D</td>
<td>Cultivated surface less than 20%</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Source: ENA 2013, INE
Finally, the sampling frame remained built with characteristics detailed in Table 3.

### Table 3. Characteristics of the area sampling frame

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Description</th>
<th>Surface km²</th>
<th>Segment size (ha)</th>
<th>Number of segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Cultivated surface greater than 60% and large fields</td>
<td>20 086</td>
<td>25.00</td>
<td>80 344</td>
</tr>
<tr>
<td>B</td>
<td>Cultivated surface greater than 60% and small fields</td>
<td>2 444</td>
<td>6.25</td>
<td>39 104</td>
</tr>
<tr>
<td>C</td>
<td>Cultivated surface between 20% and 60%</td>
<td>19 641</td>
<td>50.00</td>
<td>39 282</td>
</tr>
<tr>
<td>D</td>
<td>Cultivated surface less than 20%</td>
<td>31 370</td>
<td>100.00</td>
<td>31 370</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>Total</strong></td>
<td><strong>73 541</strong></td>
<td><strong>190 100</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Source: ENA 2013, INE*

**Sample design**

A sample of 1 500 segments was used, with segments distributed to the strata according to Neyman’s criteria. Sample selection was done through a random stratified process considering five replicates. The summary of the sample characteristic is shown below.
<table>
<thead>
<tr>
<th>Stratum H Name</th>
<th>Description</th>
<th>Surface (ha)</th>
<th>Stratum size (ha)</th>
<th>Segment number s $N_h$</th>
<th>Copies $R$</th>
<th>Zones $m_h = \frac{n_h}{r}$</th>
<th>Sample $n_h$</th>
<th>Sample by zone $N_{hj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A</td>
<td>Cultivated surface &gt; 60% and large fields</td>
<td>2,008,600</td>
<td>25.00</td>
<td>80,344</td>
<td>5</td>
<td>104</td>
<td>520</td>
<td>56 of 773 (48 of 772)</td>
</tr>
<tr>
<td>2 B</td>
<td>Cultivated surface &gt; 60% and small fields</td>
<td>244,400</td>
<td>6.25</td>
<td>39,104</td>
<td>5</td>
<td>13</td>
<td>65</td>
<td>3008</td>
</tr>
<tr>
<td>3 C</td>
<td>Cultivated surface between 20% and 60%</td>
<td>1,964,100</td>
<td>50.00</td>
<td>39,282</td>
<td>5</td>
<td>102</td>
<td>510</td>
<td>12 of 386 (90 of 385)</td>
</tr>
<tr>
<td>4 D</td>
<td>Cultivated surface less than 20%</td>
<td>3,137,000</td>
<td>100.00</td>
<td>31,370</td>
<td>5</td>
<td>81</td>
<td>405</td>
<td>23 of 388 (58 of 387)</td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td><strong>7,354,100</strong></td>
<td><strong>190,100</strong></td>
<td></td>
<td><strong>300</strong></td>
<td></td>
<td></td>
<td><strong>1500</strong></td>
<td></td>
</tr>
</tbody>
</table>
**Estimators used**

The estimate at the national level of the surface with a determined crop “c”, was obtained by:

\[ \hat{Y}_c = \frac{1}{5} \sum_{h=1}^{4} \sum_{j=1}^{m_h} N_{hj} \bar{y}_{hj} \]

Where:

\( \hat{Y}_c \) = *is the estimate for the grand surface with the crop c*

\( N_{hj} \) = *is the number of segments that build a zone j within the stratum h.*

\( y_{hji} \) = *is the crop surface c in the segment i of the zone j within the stratum h*

\( \bar{y}_{hj} \) = *is the average between segments of the crop surface c in the zone j within the stratum h*

\( m_h \) = *is the number of zones that build the stratum h*

The variance of the grand estimated surface was estimated the following way (because the number of replicates, \( n_{hj} \), is always equal to 5, according to the sample design):

\[ \hat{V}(\hat{Y}_c) = \frac{1}{5} \sum_{h=1}^{4} \sum_{j=1}^{m_h} N_{hj} (N_{hj} - 5) \hat{s}_{hj}^2 \]

However, due to non-coverage problems, sometimes the number of replicates was \( n_{hj} < 5 \), and to take into account these cases, we used the more general formula:

\[ \hat{V}(\hat{Y}_c) = \sum_{h=1}^{4} \sum_{j=1}^{m_h} N_{hj} (N_{hj} - n_{hj}) \hat{s}_{hj}^2 / n_{hj} \]

\[ \hat{s}_{hj}^2 = \frac{\sum_{i=1}^{n_{hj}} (y_{hji} - \bar{y}_{hj})^2}{n_{hj} - 1} \]
Where:

\[ \hat{V}(\hat{Y}_c) = \text{Estimated variance of the estimate surface} \]

\[ \hat{S}_{hj}^2 = \text{is the variance between segments of the crop surface c in the zone j} \]

within the stratum h

\[ \hat{S}_{hj}^2 = \frac{\sum_{i=1}^{n} (y_{hji} - \bar{y}_{hj})^2}{n_{hj} - 1} \]

Field operation

A map, by department, with the location of the sample segments was prepared.

For each segment, two images were edited using ArcGis: one with the localization scaled map 1:40 000 and the other with an orthophoto background for a location and identification field guide. Both images have overlapped a polygon of each of the scaled segments of 1:3 000 for stratum A; 1:1 000 for strata B and C; and 1:4 000 for stratum D.
Alta Verapaz Department - San Pedro Carchá Township

Each printed segment was covered with self-adherent plastic and acetate, on which the observed polygons were drawn during the field work.

Quiché Department - Chichicastenango Township
The pollsters got to the selected segments using a global position system (GPS) for their location. They first updated the soil use within each of the segments, drawing the polygons of each field on the aerial photograph and classifying them according to use codes, and then filled out the corresponding questionnaire. For this phase (two months) 60 pollsters, 15 supervisors and 4 instructors were hired.

Number and percentage of the segments planned in the sample and percent coverage are shown below.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Planned</th>
<th>Done</th>
<th>No response</th>
<th>Coverage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>520</td>
<td>511</td>
<td>9</td>
<td>98</td>
</tr>
<tr>
<td>B</td>
<td>65</td>
<td>63</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td>C</td>
<td>510</td>
<td>490</td>
<td>20</td>
<td>96</td>
</tr>
<tr>
<td>D</td>
<td>405</td>
<td>367</td>
<td>38</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>1500</td>
<td>1431</td>
<td>69</td>
<td>95</td>
</tr>
</tbody>
</table>
The following chart shows the tools with which the data in the fields were measured.
Data processing

A process of criticism and codification was implemented prior to digitizing the drawn polygons over the orthophoto, generating a geographic database in a shapefile format to which a consistency analysis was done, and each field was linked with the information of the corresponding cards.

Digitizing survey cards data

The data obtained from the survey cards, after its critical review and codification, was digitized using a program in CsPro.
## Results obtained and precision of estimates

Table 5. Surface estimates of soil use, August to November 2013

<table>
<thead>
<tr>
<th>Crop or soil use</th>
<th>Surface at national level (ha)</th>
<th>Typical error</th>
<th>95% confidence interval</th>
<th>Variation coefficient (%)</th>
<th>Design effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inferior</td>
<td>Superior</td>
<td></td>
</tr>
<tr>
<td>Sampling frame total</td>
<td>7,354,115.11</td>
<td>18.21</td>
<td>7,354,079.39</td>
<td>7,354,150.84</td>
<td>0.00025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00000</td>
</tr>
<tr>
<td>Agricultural (arable soil)</td>
<td>3,524,213.83</td>
<td>60,172.43</td>
<td>3,406,151.36</td>
<td>3,642,276.30</td>
<td>1.70</td>
</tr>
<tr>
<td>Annual crops</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>Rice</td>
<td>889,711.31</td>
<td>33,882.55</td>
<td>823,231.39</td>
<td>956,191.23</td>
<td>3.80</td>
</tr>
<tr>
<td>Beans</td>
<td>184,181.54</td>
<td>15,822.21</td>
<td>153,137.26</td>
<td>215,225.81</td>
<td>8.60</td>
</tr>
<tr>
<td>Corn</td>
<td>803,989.65</td>
<td>32,073.98</td>
<td>741,058.29</td>
<td>866,921.02</td>
<td>4.00</td>
</tr>
<tr>
<td>Potato</td>
<td>3,803.53</td>
<td>1,224.31</td>
<td>1,401.34</td>
<td>6,205.72</td>
<td>32.20</td>
</tr>
<tr>
<td>Vegetables</td>
<td>46,888.42</td>
<td>5,820.87</td>
<td>35,467.47</td>
<td>58,309.36</td>
<td>12.41</td>
</tr>
<tr>
<td>Other annual crops</td>
<td>19,849.36</td>
<td>3,582.94</td>
<td>12,819.38</td>
<td>26,879.34</td>
<td>18.05</td>
</tr>
<tr>
<td>Permanent crops</td>
<td>1,033,132.82</td>
<td>47,102.28</td>
<td>940,714.89</td>
<td>1,125,550.74</td>
<td>4.56</td>
</tr>
<tr>
<td>Coffee</td>
<td>363,755.07</td>
<td>28,032.45</td>
<td>308,753.46</td>
<td>418,756.67</td>
<td>7.70</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>260,841.18</td>
<td>23,356.36</td>
<td>215,014.39</td>
<td>306,667.98</td>
<td>9.00</td>
</tr>
<tr>
<td>Cardamom</td>
<td>82,230.58</td>
<td>21,637.14</td>
<td>39,777.01</td>
<td>124,684.15</td>
<td>26.30</td>
</tr>
<tr>
<td>Rubber</td>
<td>97,358.01</td>
<td>16,696.29</td>
<td>64,598.72</td>
<td>130,117.29</td>
<td>17.10</td>
</tr>
<tr>
<td>African palm tree</td>
<td>115,001.77</td>
<td>19,126.12</td>
<td>77,474.99</td>
<td>152,528.54</td>
<td>16.60</td>
</tr>
<tr>
<td>Other permanent crops</td>
<td>152,550.55</td>
<td>17,011.76</td>
<td>119,172.30</td>
<td>185,928.80</td>
<td>11.15</td>
</tr>
<tr>
<td>Agricultural lands without crops</td>
<td>113,213.89</td>
<td>10,981.48</td>
<td>91,667.46</td>
<td>134,760.32</td>
<td>9.70</td>
</tr>
<tr>
<td>Shrubs and bushes</td>
<td>669,213.14</td>
<td>40,136.04</td>
<td>590,463.44</td>
<td>747,962.84</td>
<td>6.00</td>
</tr>
<tr>
<td>Grass</td>
<td>1,685,170.15</td>
<td>57,028.63</td>
<td>1,573,276.03</td>
<td>1,797,064.27</td>
<td>3.38</td>
</tr>
</tbody>
</table>

59 annual and 45 permanent crops were studied, and the results are presented in the chart below.
Table 7. Surface results of soil uses at the national level

<table>
<thead>
<tr>
<th>Crop or soil use</th>
<th>Surface at national level (ha)</th>
<th>Typical error</th>
<th>Variation coefficient (%)</th>
<th>Design effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sesame</td>
<td>7,684.85</td>
<td>2,324.97</td>
<td>0.30</td>
<td>0.85</td>
</tr>
<tr>
<td>Rice</td>
<td>14,348.48</td>
<td>4,644.92</td>
<td>0.32</td>
<td>0.83</td>
</tr>
<tr>
<td>Sweet pea</td>
<td>1,708.92</td>
<td>735.72</td>
<td>0.43</td>
<td>1.62</td>
</tr>
<tr>
<td>Chinese pea</td>
<td>2,015.96</td>
<td>896.17</td>
<td>0.44</td>
<td>0.92</td>
</tr>
<tr>
<td>Oat</td>
<td>1,989.67</td>
<td>685.83</td>
<td>0.34</td>
<td>1.39</td>
</tr>
<tr>
<td>Broccoli</td>
<td>1,142.15</td>
<td>539.42</td>
<td>0.47</td>
<td>0.64</td>
</tr>
<tr>
<td>Green pea</td>
<td>2,650.48</td>
<td>1,644.55</td>
<td>0.62</td>
<td>0.75</td>
</tr>
<tr>
<td>Bean (monoculture)</td>
<td>18,405.13</td>
<td>3,137.07</td>
<td>0.17</td>
<td>0.82</td>
</tr>
<tr>
<td>Sorghum</td>
<td>1,228.10</td>
<td>743.19</td>
<td>0.61</td>
<td>1.10</td>
</tr>
<tr>
<td>Ensilage corn</td>
<td>1,002.97</td>
<td>658.47</td>
<td>0.66</td>
<td>0.59</td>
</tr>
<tr>
<td>Ensilage corn (monoculture)</td>
<td>2,956.20</td>
<td>1,647.63</td>
<td>0.56</td>
<td>0.59</td>
</tr>
<tr>
<td>Corn (monoculture)</td>
<td>529,774.32</td>
<td>26,941.56</td>
<td>0.05</td>
<td>0.64</td>
</tr>
<tr>
<td>Peanut</td>
<td>5,376.83</td>
<td>2,373.65</td>
<td>0.44</td>
<td>0.61</td>
</tr>
<tr>
<td>Potato</td>
<td>3,732.84</td>
<td>1,223.13</td>
<td>0.33</td>
<td>1.57</td>
</tr>
<tr>
<td>Cabagge</td>
<td>2,016.83</td>
<td>702.67</td>
<td>0.35</td>
<td>2.34</td>
</tr>
<tr>
<td>Carrot</td>
<td>934.28</td>
<td>325.45</td>
<td>0.35</td>
<td>2.59</td>
</tr>
<tr>
<td>Achiote</td>
<td>5,239.93</td>
<td>1,587.06</td>
<td>0.30</td>
<td>0.71</td>
</tr>
<tr>
<td>Avocado</td>
<td>5,312.90</td>
<td>3,014.11</td>
<td>0.57</td>
<td>2.18</td>
</tr>
<tr>
<td>Banana</td>
<td>26,997.95</td>
<td>8,539.12</td>
<td>0.32</td>
<td>1.06</td>
</tr>
<tr>
<td>Cacao</td>
<td>2,992.64</td>
<td>2,394.75</td>
<td>0.80</td>
<td>0.59</td>
</tr>
<tr>
<td>Coffee (monoculture)</td>
<td>250,406.55</td>
<td>23,382.57</td>
<td>0.09</td>
<td>0.79</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>260,841.18</td>
<td>23,356.36</td>
<td>0.09</td>
<td>0.81</td>
</tr>
<tr>
<td>Cardamom</td>
<td>82,230.58</td>
<td>21,637.14</td>
<td>0.26</td>
<td>1.08</td>
</tr>
<tr>
<td>Rubber (monoculture)</td>
<td>85,544.85</td>
<td>15,873.43</td>
<td>0.19</td>
<td>0.88</td>
</tr>
<tr>
<td>Jocote of other kind</td>
<td>1,597.71</td>
<td>847.85</td>
<td>0.53</td>
<td>0.89</td>
</tr>
<tr>
<td>Lemon</td>
<td>3,876.95</td>
<td>1,635.07</td>
<td>0.42</td>
<td>0.68</td>
</tr>
<tr>
<td>Macadamia</td>
<td>10,440.14</td>
<td>4,986.34</td>
<td>0.48</td>
<td>0.83</td>
</tr>
<tr>
<td>Mango</td>
<td>3,745.61</td>
<td>1,579.20</td>
<td>0.42</td>
<td>0.94</td>
</tr>
<tr>
<td>Apple</td>
<td>1,613.03</td>
<td>1,488.63</td>
<td>0.92</td>
<td>0.58</td>
</tr>
<tr>
<td>Pineapple</td>
<td>3,859.00</td>
<td>2,599.30</td>
<td>0.67</td>
<td>1.14</td>
</tr>
<tr>
<td>Platano</td>
<td>9,319.82</td>
<td>4,064.65</td>
<td>0.44</td>
<td>1.16</td>
</tr>
</tbody>
</table>
Resources mobilized and other cost

The implemented budget to carry out the survey was approximately of USD 300,000.00.

Issues and challenges encountered during implementation and how they were addressed

Given the innovative nature of this methodology in Guatemala, several issues were found:

- Lack of trained and permanent personnel.
- Access roads in very bad condition, and many times it was not practicable to conduct field work during the rainy season.
- Lack of coverage because of inability to get to some segments (i.e. in situations where there were no roads).
- Lack of coverage due to the lack of collaboration from the population, high tensions within the country due to strip mining, hydroelectric plants, extortions, taxes, and other issues.
- Lack of equipment (mainly GPSs and vehicles).
- Communication difficulties because at least 24 languages are spoken in the country.
- The high ecologic variability allows farming at various periods and at very disperse dates throughout the year; therefore, one visit is not enough to properly estimate cultivated area and other variables.
- In the field, the main problems were:
  - Lack of clarity in the layout of polygons.
  - Inconsistency in the number of fields of the card and the number of fields (polygons) in the image.
  - Different codes used for the same soil use.
  - Illegible cards.
  - Number of duplicated fields, fields without numbers or with row-use codes.
  - The use code in the picture was different from that in the consigned card.
NEPAL

By Ambika Bashyal

Introduction and background

Nepal is a landlocked country that is bordered on the south and west by India and by China to the north. The average east–west length of the country is 885 km, and the average north–south breadth is 193 km. The country’s elevation varies from a low of 90 m above sea level in the southeast, to 8,848 m (the highest peak in the world: Everest). As a result of this diversity in topography, the country’s weather and climate is also very diverse, ranging from tropical to artic.

Administratively, the country is divided into 5 development regions and 75 districts. Village development committees and municipalities are the lowest administrative units within districts.

Ecologically, the country is divided into three north–south regions: mountains, hills and the plains called tara. From an agricultural point of view, the mountainous regions are suitable for livestock raising, the hills for horticulture, and the tarai for cereal crops such as rice, wheat and maize. Based on the area of the 75 districts, these three regions — mountains, hills and the tarai — constitute 35%, 42% and 23%, respectively, of the country’s total land area of 147,181 km².

The agriculture sector is crucial for the economic development of Nepal. More than one-third of Nepal’s gross domestic product (GDP) comes from agricultural activities, and these absorb about 61% of the labor force. Agriculture will continue to be the most important sector of the national economy at least until the foreseeable future. The country has suitable conditions for farming not only a wide variety of crops but also for raising livestock.

The Government of Nepal has accorded top priority to agricultural development in all of its five-year plans. Since 1995, Nepal has been implementing the 20-year-long Agricultural Perspective Plan (APP) to overcome the problem of food insecurity and for alleviating poverty. Although top priority is given to the agriculture sector, the food situation in the country is much below requirement. According to the National Sample Census of Agriculture 2011/2012, the number of agricultural holdings in the country is 3.8 million while the total area operated is 2.5 million ha.
Agricultural data

The structural agricultural statistical data have been collected through decennial agricultural censuses undertaken by the Central Bureau of Statistics (CBS) since 1961/62. Current agricultural statistics (area under various crops, livestock and their products), however, are being collected through different surveys, observations and official records of the Ministry of Agriculture Development.

Agricultural census

The Central Bureau of Statistics is responsible for carrying out decennial agricultural censuses, which are important sources of statistics regarding the structure of agriculture. Nepal has been conducting agricultural censuses under the recommendations and framework of the Food and Agriculture Organization of the United Nations (FAO) World Programme for the census of agriculture since 1961/62. Agricultural censuses conducted so far in Nepal are for the years: 1961/62, 1971/72, 1981/82, 1991/92, 2001/02 and 2011/12.

All of these censuses were conducted on a sampling basis.

The basic information needed for frame building for the agricultural census is the number of agriculture holdings, which is based on the information collected during the population census at the time of household listing. This is the ultimate sampling unit of the agricultural census.

Agriculture holding

Whether an agricultural census is based on a complete enumeration or on a sample, the aim is to obtain information on various characteristics of “operational holdings”\(^{20}\). Hence, in an agricultural census, operational holdings constitute the population, and the frame, which is used in the data collection phase, is defined as the list of operational holdings. An agricultural holding is an economic unit of agricultural production under a single management, and comprising all livestock and poultry kept, and all land used wholly or partly for agricultural production purposes.

---

\(^{20}\) A holder is a person who manages the agricultural operations within a holding, makes decisions about the utilization of agricultural resources and the distribution of produce, and who bears the technical and financial responsibility of the operation. The operational holding is the operating unit concept that includes the holder who may be full owner, part owner, or tenant irrespective of the ownership of the land or livestock.
A holding is considered to be an agricultural unit if it satisfies any one of the following conditions:

- has an area under crops in the hills and mountains that is greater than or equal to 4 ana (0.01272 ha) or greater than or equal to 8 dhur (0.01355 ha) in the tarai; or
- keeping 1 or more head of cattle or buffaloes; or
- keeping 5 or more head of sheep or goats; or
- keeping 20 or more poultry.

A household with livestock but no land is not considered a holding if the household does not use the livestock for agricultural purposes (e.g. a livestock trader). A holding includes all land operated within the district, whether it is owned by the holding or not. The holding’s land may consist of one or more parcels located in one or more separate areas. The holding is generally the same as a household. Public grazing land does not constitute a holding. Land owned jointly by more than one person for grazing or other purposes is included. Holdings are defined in terms of their land and livestock holdings on the day of the census enumeration.

In 2011/12, the CBS conducted a National Sample Census of Agriculture (2011/12 NSCA) based on experiences from previous censuses. In order to meet the growing needs of users, the 2011/12 NSCA had the following objectives:

- collect data from around 130 000 agricultural holdings;
- provide detailed tabulation and analyses;
- publish, at the district level, data on structure and characteristics of holdings (e.g. size); agricultural land use, tenure and fragmentation; area planted in crops; number of livestock; and other farm practices;
- provide benchmark data on agriculture in order to improve the reliability of estimates from the current agricultural surveys; and
- provide sample frames for the current and future agricultural surveys.

Main data items and variables

The 2011/12 NSCA adopted most of FAO’s recommendations regarding data collection under the following main headings: Identification, Holder characteristics, Demographic and other characteristics, Employment, Land and
water, Crops, Livestock and poultry, Agricultural inputs, Agricultural machinery and equipment, Non-residential buildings, Agricultural credit, Forest, Fishery and Ancillary activities on the holding.

**Methodology for building the sampling frame and the survey design**

**Master sampling frame**

In Nepal, a master sampling frame was not prepared for the agricultural census and for survey purposes. However, the information required for the master sampling frame is available, both from the listing of the population census and the agricultural census. The table of sample allocation (Annex 1) can be considered as the master sampling frame for the agricultural surveys. The CBS has not yet initiated any preliminary work for the purpose of preparing the area sampling frame.

**Sampling design**

The sampling design adopted in the 2011/12 NSCA was a stratified, two-stage sampling with the district as strata, an individual ward\(^{21}\) or a sub-ward or a group of contiguous wards as the primary sampling units (PSUs), and agricultural holdings as the ultimate sampling units (USUs) for district-wide publication of the data. Each of the 75 districts was considered a stratum.

The first-stage selection was done using probability proportional to size systematic sampling, with the number of holdings in PSUs as measures of size. Selection of agricultural holdings at the second stage was done using equal probability systematic sampling. The overall sampling design was self-weighted within each district, provided the number of households in every enumeration area (EA) at the time of listing and in the population census is the same. Differences in these numbers, if any, were easily taken care of at the selection stage of the number of agricultural holdings by adjusting the holdings from 20 to 30. The sample selection was done in such a way that all holdings in a given district had the same chance of selection in the sample. Manang District was completely enumerated because it had only a small number of holdings.

---

\(^{21}\) A ward is the smallest administrative unit of the government and has a clearly-defined boundary and is mostly used as the enumeration area for census and survey purposes. The number of wards is fixed at 9 for the Village Development Committees in rural areas, and ranges from 9 to 35 in municipalities in urban areas.
Sources of sampling frames

Two types of sampling frames were used in the sampling design for the 2011/12 NSCA: one for selecting PSUs, and the other for selecting holdings. The sampling frame was prepared from the list of EAs for PSUs from the 2011 population census. The frame contained a listing of wards and blocks, along with the number of houses, households and agricultural households (holdings).

In order to prepare a frame for the agricultural holdings, a systematic listing of agricultural households was carried out in each selected PSU.

Frame for primary sampling units

The “Household Listing Form” of the 2011 population census contained the following information on agriculture: the number and area of holdings, and the number of livestock and poultry. The agriculture module in the 2011 population census of Nepal was as follows.

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Does your household operate any land for agricultural purpose?</td>
<td>1 Yes, 2 No</td>
</tr>
<tr>
<td>11. If yes, the total area of the land in</td>
<td>1) Bigha, 2) Katha, 3) Dhur</td>
</tr>
<tr>
<td>12. If yes, the total number of livestock or poultry for agriculture</td>
<td>1) Yes, 2) No</td>
</tr>
</tbody>
</table>
The 2011 population census had prepared EA maps for 44 municipalities in the country and for village development committees (VDCs) in 12 district headquarters. Further, EA (block) maps were prepared for a few particularly large VDCs. These EA maps were used during the fieldwork of the 2011/12 agricultural census.

According to the 2011 census, a ward within a VDC or municipality contained, on average, about 150 agricultural households, but there was wide variation among wards. On the whole, there were almost 1 600 wards with very small numbers of agricultural households (less than 25). These small wards needed to be combined with neighboring wards in the same VDC so as to ensure enough agricultural households in each PSU for the final stage of selection. In contrast, for some very large wards, some kind of segmentation of wards was necessary.

**Combination of small wards.** Because the average number of agricultural households (holdings) to be enumerated in each ward was set at 25, there would be a problem if the ward contained less than 25 holdings. To avoid this problem, a small ward was grouped with one or more neighboring ward(s) of the same VDC so that the combined total exceeded 25 agricultural households. Whenever such a combined PSU of wards was selected, all the wards were taken for the second stage listing, which was the listing of holdings.

**Segmentation of large wards.** In the case of very large wards, it was necessary to use segmentation for making sizable PSUs. For such wards, the block maps prepared for the 2011 population census were used.

---

22 The number of holdings for each primary sampling unit is a minimum of 20 and a maximum of 30.
Sample size

The tentative sample size for the 2011/12 agricultural census was estimated at 130,000 agricultural households (holdings) selected from 5,200 PSUs across the country. The number of PSUs to be allocated for each district was determined on the basis of the total area under nine major crops (paddy, wheat, maize, millet, barley, sugarcane, oilseed potato and vegetables) in the district ($X_0$). The average area of the nine major crops for the last three years published by the Ministry of Agriculture Development was the measure of size for the allocation. The number of PSUs ($n_d$) to be selected from a district was based on the “compromise power” allocation:

$$n_d \propto (X_0)^{0.4}$$

In the 2011/12 agriculture census, 5,200 sample holdings were allocated to the 75 districts, following the compromise power allocation method. Samples were allocated to different districts in proportion to $X^{0.4}$ where $X$ was the size measure.

Selection of primary sampling units

The sketches and maps used in the 2011 population census were used for preparing the frames of PSUs by merging nearby wards (or sub-wards) in cases where the number of agricultural holdings was less than 25. As previously mentioned, around 1,600 wards were found to have less than 25 holdings.

Information on agricultural households was transcribed from 36,022 wards of 3,960 VDCs and municipalities taken together. From these wards, 36,257 PSUs were constructed after combining small wards and segmenting very large ones. Out of 36,257 PSUs thus formed, 1,029 were combined PSUs and 396 PSUs were segmented PSUs.

Of the total PSUs formed by combining two or more wards, the highest proportion was in the hills (47%) followed by those in the mountains (34%). And, of the total PSUs formed by the segmentation of large wards, the highest proportion was in the hills (54%) followed by those in the tarai (42%).

Prior to selection, the data for each district was tabulated so that the PSUs were arranged by VDC, and alternately in ascending and descending order of the size of the agricultural holdings of PSUs. This was to ensure implicit stratification and to increase the efficiency of the probability proportion to size systematic sampling. After arranging the data in this way, the selection was done using the IBM SPSS 20 (Statistical Package for Social Sciences) software package.
**List of selected PSUs.** After selecting the required number of PSUs in each district, as determined by the method of compromise power allocation, the list of selected PSUs to be used in the field was prepared. The selected list of the PSUs is presented in Annex 1.

**Selection of the ultimate sampling units**

The frame for the selection of holdings was prepared during the listing of agricultural households. In each of the selected PSUs, a list of holders was prepared. Within the selected PSUs, all of the operational agricultural holdings were listed and classified into the following four categories:

1. holdings operating less than 1 bigha (6772.63 m² )/10 ropanis (1 ropani= 508.72 m² ) of land,
2. holdings operating equal to or more than 1 bigha/10 ropanis of land but less than 3 bighas/20 ropanis of land,
3. holdings operating equal to or more than 3 bighas/20 ropanis of land, and
4. holdings having no land but keeping one or more productive animals and/or keeping 20 or more poultry birds.

Finally, the 25 holdings for the enumeration were selected systematically with a random start. Immediately after the listing operation, interviews were conducted. The implicit stratification has partial advantages of stratification without making the strata explicitly, and can be used for independent selections within each stratum.

**Formulas for estimators**

All parameters were first estimated at the district level, and the parameters of the development regions, ecological belts and the national estimates, were obtained by aggregating the data of the respective districts. The formulas used to calculate parameter estimates, standard error and design effect are presented in Annex 2.
Results obtained and precision of estimates

The mean and total of the major characteristics — total area of holdings, area of temporary crops, area under paddy, area under wheat, area under maize, number of cattle, number of buffaloes, number of chickens and farm population — were calculated and presented in national-, regional- and district-level reports, which contain 32 output tables with explanatory text. A few of the major findings are given in the following three tables.

Table 1. Area and fragmentation of holdings, Nepal, 1961/62–2011/12

<table>
<thead>
<tr>
<th>Category</th>
<th>Census year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of holdings (’000)</td>
<td></td>
</tr>
<tr>
<td>Holdings with land</td>
<td>1,518.0</td>
</tr>
<tr>
<td>Holdings with no land</td>
<td>22.0</td>
</tr>
<tr>
<td>Total holdings (’000)</td>
<td>1,540.0</td>
</tr>
<tr>
<td>Holdings with land</td>
<td>1,685.4</td>
</tr>
<tr>
<td>Area of holdings (’000 ha)</td>
<td>1.1</td>
</tr>
<tr>
<td>Average parcel per holding</td>
<td>6.8</td>
</tr>
<tr>
<td>Average parcel size (ha)</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 2. Characteristics of selected holdings in the different development regions and ecological belts, 2011/12.

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of holdings (’000)</th>
<th>Area of holdings (’000 ha)</th>
<th>Average holding size (ha)</th>
<th>Paddy Holdings (’000)</th>
<th>Area (’000)</th>
<th>Maize Holdings (’000)</th>
<th>Area (’000)</th>
<th>Wheat Holdings (’000)</th>
<th>Area (’000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern</td>
<td>894.9</td>
<td>755.2</td>
<td>0.8</td>
<td>658.9</td>
<td>482.0</td>
<td>311.6</td>
<td>157.6</td>
<td>515.5</td>
<td>190.5</td>
</tr>
<tr>
<td>Central</td>
<td>1153.9</td>
<td>716.9</td>
<td>0.6</td>
<td>942.8</td>
<td>447.3</td>
<td>582.0</td>
<td>229.3</td>
<td>645.9</td>
<td>206.6</td>
</tr>
<tr>
<td>Western</td>
<td>797.3</td>
<td>482.6</td>
<td>0.6</td>
<td>594.5</td>
<td>244.8</td>
<td>349.7</td>
<td>97.7</td>
<td>566.0</td>
<td>136.4</td>
</tr>
<tr>
<td>Mid-western</td>
<td>575.1</td>
<td>353.6</td>
<td>0.6</td>
<td>381.8</td>
<td>153.0</td>
<td>430.5</td>
<td>139.6</td>
<td>450.8</td>
<td>108.9</td>
</tr>
<tr>
<td>Far western</td>
<td>409.9</td>
<td>217.4</td>
<td>0.5</td>
<td>376.8</td>
<td>128.8</td>
<td>360.2</td>
<td>125.2</td>
<td>236.2</td>
<td>31.4</td>
</tr>
<tr>
<td>Ecological belts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mountain</td>
<td>327.5</td>
<td>213.9</td>
<td>0.7</td>
<td>238.7</td>
<td>61.7</td>
<td>217.1</td>
<td>56.7</td>
<td>277.3</td>
<td>74.2</td>
</tr>
<tr>
<td>Hill</td>
<td>1729.7</td>
<td>986.1</td>
<td>0.6</td>
<td>1154.7</td>
<td>309.5</td>
<td>805.2</td>
<td>185.9</td>
<td>1577.4</td>
<td>453.1</td>
</tr>
<tr>
<td>Terai</td>
<td>1773.9</td>
<td>1325.6</td>
<td>0.8</td>
<td>1561.3</td>
<td>1084.8</td>
<td>1011.8</td>
<td>506.8</td>
<td>559.7</td>
<td>146.4</td>
</tr>
<tr>
<td>Nepal</td>
<td>3831.1</td>
<td>2525.6</td>
<td>0.7</td>
<td>2954.8</td>
<td>1456.0</td>
<td>2034.1</td>
<td>749.4</td>
<td>2414.4</td>
<td>673.7</td>
</tr>
</tbody>
</table>
Table 3. Holdings with livestock and livestock number, Nepal 1991/92–2011/12

<table>
<thead>
<tr>
<th>Livestock type</th>
<th>Holdings ('000)</th>
<th>Livestock population ('000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Livestock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle</td>
<td>2067.1</td>
<td>2297.1</td>
</tr>
<tr>
<td>Chaunri</td>
<td>9.0</td>
<td>14.2</td>
</tr>
<tr>
<td>Buffaloes</td>
<td>1307.8</td>
<td>1586.8</td>
</tr>
<tr>
<td>Goats</td>
<td>1382.8</td>
<td>1686.4</td>
</tr>
<tr>
<td>Sheep</td>
<td>92.7</td>
<td>67.7</td>
</tr>
<tr>
<td>Pigs</td>
<td>267.5</td>
<td>227.8</td>
</tr>
<tr>
<td>Poultry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chickens</td>
<td>1400.4</td>
<td>1594.4</td>
</tr>
<tr>
<td>Ducks</td>
<td>92.6</td>
<td>110.1</td>
</tr>
<tr>
<td>Pigeons</td>
<td>215.8</td>
<td>202.1</td>
</tr>
<tr>
<td><strong>Total holdings</strong></td>
<td><strong>2736.1</strong></td>
<td><strong>3364.1</strong></td>
</tr>
</tbody>
</table>

Sampling errors

For the first time in Nepal in 1994, an assessment of data quality was carried out with the technical assistance of the FAO, and was based on the data of the 1991/92 agricultural census. A brief overview of the sampling errors and their measurements and interpretations were provided, together with an overall assessment of the reliability of the census data from a sampling error point of view. Tables were provided, showing a number of sample design parameters for major characteristics, which were utilized to finalize the sample design of the 2011/12 agricultural census.

Due to some limitations, the assessment of the reliability of the 1991/92 census data could not be carried over into the 2001/02 agricultural census. In the 2011/12 census, some preliminary exercises were done to calculate the values of some of the indicators identified in 1994 with the assistance of an FAO technical report on the reliability of data.

Standard errors have been calculated for: total area of holdings; area under temporary crops; area under paddy and maize; area under wheat; area under potatoes; area of irrigated land; total paddy production; total maize production; total wheat production; total potatoes production; and farm population. The SPSS version 20 software was used for these calculations. The standard errors...
of the exercise for the ecological belts and development regions are presented in Annex 3, and the sample design parameters of the exercise for the districts are presented in Annex 4.

**Mobilization of resources and other costs**

To carry out any large-scale agricultural census, qualified labor, powerful computers for data processing, and other resources are needed. In order to select the representative sample, a sample design expert and a computer software programming expert for data processing are also required. The required human resources were managed from two sources. The main source of census staff was the pool of regular staff working at the CBS central office and at the statistical offices in the districts. The other group of census staff was short-term contract recruits. There were 1,600 enumerators, 550 supervisors, 75 census officers and 350 support staff involved in the field work of the 2011/12 agricultural census, excluding the census core team at the head office.

The cost of the 2011/12 NSCA was about USD 1.43 million and the salary of regular staff was not included in this cost. The tabular breakdown of cost is shown in Table 4.
Table 4. Cost for the 2011/12 National Sample Census of Agriculture

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>Cost</th>
<th>Budget in NPR</th>
<th>Budget in USD*</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pilot agriculture census</td>
<td>8,000,000.00</td>
<td>80,000.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Machinery, equipment</td>
<td>1,500,000.00</td>
<td>15,000.00</td>
<td>Computers, printers, photocopiers, scanners, cameras</td>
</tr>
<tr>
<td>3</td>
<td>Consultancies and other services</td>
<td>2,200,000.00</td>
<td>22,000.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Frame construct for agriculture census from population census</td>
<td>1,500,000.00</td>
<td>15,000.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>household listing form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Printing</td>
<td>5,500,000.00</td>
<td>55,000.00</td>
<td>Questionnaire forms, manuals, control forms, administrative forms, financial administration forms, and other materials</td>
</tr>
<tr>
<td>6</td>
<td>Logistics</td>
<td>5,000,000.00</td>
<td>50,000.00</td>
<td>Enumerators' bags, clipboards, calculators, torch lights, backpacks, black and red dot pens</td>
</tr>
<tr>
<td>7</td>
<td>Training</td>
<td>6,000,000.00</td>
<td>60,000.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Transportation</td>
<td>4,000,000.00</td>
<td>40,000.00</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Media campaigning</td>
<td>2,500,000.00</td>
<td>25,000.00</td>
<td>AM/FM radio, television, newspapers, posters, pamphlets, leaflets, folders, banners, and other materials</td>
</tr>
<tr>
<td>10</td>
<td>Field operation</td>
<td>90,000,000.00</td>
<td>900,000.00</td>
<td>Including central-level supervision</td>
</tr>
<tr>
<td>11</td>
<td>Data processing</td>
<td>6,000,000.00</td>
<td>60,000.00</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Report printing</td>
<td>6,000,000.00</td>
<td>60,000.00</td>
<td>85 reports of different levels</td>
</tr>
<tr>
<td>13</td>
<td>Results dissemination at different levels</td>
<td>3,200,000.00</td>
<td>32,000.00</td>
<td>At the national and district level</td>
</tr>
<tr>
<td>14</td>
<td>Miscellaneous</td>
<td>2,000,000.00</td>
<td>20,000.00</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Total</td>
<td>143,400,000.00</td>
<td>1,434,000.00</td>
<td></td>
</tr>
</tbody>
</table>

*1 USD = 100 NPR
Census data collection

The primary reference period for the data collected during the census was the calendar year 2011. Land area and livestock data were collected on the day of enumeration.

For data collection, a district census office was established in each of the country’s 75 districts for three months. The district census officers were responsible for census field operations and had an overall managerial role in their respective districts. Field staff were recruited and trained within the first month of establishing an agricultural census office. Field enumeration was conducted in the following two months.

Enumeration in the districts was done in two stages. First, the listing forms of the agricultural households were filled in and then the individual forms for the selected agricultural households were filled in.

Census enumeration work was conducted in two phases. In the first phase, from January to March 2012, enumeration was carried out in 59 districts in the tarai and hill areas. Enumeration for the remaining 16 mountain districts was conducted from April to June 2012 when the climate was more favorable. Data were collected in the 12-month reference period of January to December 2011.

Selection and training of field staff

To carry out a large-scale agricultural census, a large number of skilled labor is needed, and should include individuals with a knowledge of sample selection, computer skills, agricultural activities and methods of enumeration. The staff required for the census came from two sources. The main source was the regular staff working at the CBS and the statistical offices in the districts. The source for the census of labor at the district level was recruits hired on a three-month contract.

Staff at the central level were involved in the planning, implementation, monitoring, data processing, tabulation, results dissemination, and final report preparation. They also worked as the agricultural census officers and central supervisors. Staff recruited on contract were hired for only the field enumeration. There were four to six supervisors under an agricultural census officer, and one supervisor supervised the field operation of four enumerators.

In total, 2,500 people, excluding central-level core team members, made up the census team: 1,600 enumerators, 500 supervisors, 75 district agriculture census officers and 325 other staff.
Training the census field personnel was undertaken at three levels. The training for field personnel consisted of lectures with PowerPoint presentations, discussions, mock interviews, and practicing filling in the questionnaire.

**Training district census officers**

Training was organized for the 75 district census officers. All of the concepts and definitions used in the census schedules and procedures of field enumeration were thoroughly explained to the trainees. The training included a detailed study of the questionnaire and reference manual, classroom sessions, class exercise sessions, mock interviews and field practices and discussions.

**Training field supervisors**

Field supervisors were trained at the district headquarters by the district census officer and the central supervisor.

**Training enumerators**

This training was given high priority because the whole outcome of the census depended on how well they were trained. The training was organized at the respective district census office.

The aim of this training was to develop enumerators’ capacity to motivate the respondents in providing accurate and complete answers. This training focused on a thorough study of the questionnaire and the reference manual, reading maps, class exercises, mock interviews and field practices, and discussions with the help of PowerPoint presentations and flip charts.

Two more major central-level training programmes were conducted for data processing (coding, editing and consistency check, and data entry) at the CBS.

**Survey materials for field staff**

An agricultural census is a voluminous and complex task. The different stages of conducting a census (training and field work) require different types of survey materials and logistics. The questionnaire is the most important material for the census although field staff also need bags, pens, sleeping bags, torch lights, and other necessities for field work. Such materials are required not only to conduct the census smoothly, but also to improve the quality of the census. All of the necessary requirements of the enumerators for conducting their field work were provided to them before the beginning of the enumeration.
Organization of field work

The 2011/12 national agricultural census project was implemented by the agricultural census and survey section of the CBS’s Economic Statistics Division, with the coordination of other divisions and sections. It was the responsibility of the Director General of the CBS to implement the agricultural census. Under the direct control of the Director General, project coordinator and census officer, enumerators carried out the field enumeration with the supervision and direction of the field supervisor.

A district agricultural census office was established in each district to carry out the census. The district agricultural census office supervised the activities of its supervisors and enumerators. Similarly, supervisors took care of field enumeration under their assignment area. A hierarchical organization was made to successfully implement the census.

A high-level National Agriculture Census Steering Committee, chaired by the vice chair of the National Planning Commission, was formed to guide the activities of the census — from pre-planning to result disseminations — in an organized way. This committee supported by providing the policy direction in the different stages of the census activities and managed the necessary resources needed for the census and solved the problems encountered during the whole operation.

Similarly, to advise the CBS on technical aspects of the census, the National Agricultural Census Technical Committee, chaired by the Director General of the CBS was formed. This committee made significant recommendations during the various phases of census operations. At the district level, a District Census Coordination Committee was formed under the supervision of the Chief District Officer. The committee was responsible for coordinating and supervising the enumeration work in the district.

Questionnaires

In the agricultural census, the information was collected by filling in the questionnaire during the interview with the agricultural holder(s) of the selected agricultural holding(s). As such, extreme care was taken in designing the questionnaire. FAO guidelines were taken into consideration while designing the questionnaire for the 2011/12 agricultural census. Most of the questions used were carried over from the agricultural census of 2001/02, which facilitated comparison, uniformity and standardization.
The questionnaires were prepared with the potential qualification of the enumerators in mind. The tables were published as the results of the census. Due care was taken to capture intended information by properly constructing questions and pretesting of the questionnaire. Three types of questionnaires were used in census:

- Schedule 1 (Agricultural Holdings Listing Form),
- Schedule 1A (List of Selected Agricultural Holdings), and
- Schedule 2 (Questionnaire for Individual Holdings).

The questionnaire used for the data collection in the 2011/12 NSCA (Schedule2) is presented in Annex 5.

**Procedures for data collection and quality control**

The data were collected through an interview with the principal agricultural holder or any close member of the holder who could understand and answer the questionnaire. There was no plan to record the objective measurement of the area of the land or any measure of the harvest. The answers of the respondents were the sole basis of the information collected.

Some respondents were unable to provide the area of the land in the local units of ropani or bigha, or the amount of seed sown. The local units they used were converted into standard units, using the available conversion tables.

Quality control is a challenging task in every step of a survey, starting from the drafting of the questionnaire to the analysis of the data collected. In order to minimize non-sampling errors in the census, quality control measures were instituted at the various phases of the activity, from the conceptualization stage down to the analysis of the results.

Sampling error was minimized by modifying the design of the previous agriculture census. Training, supervision, control forms, verification and consistency checks were the quality control mechanisms used to enhance the accuracy of the final census results.
Some new features of the 2011/12 agricultural census

Production of major crops

Information on the production of major crops was included in the first agricultural census of Nepal in 1961/62. This information was not collected in subsequent censuses but was included in the 2011/12 census. The inclusion of major crops the production was to validate annual crop production estimates compiled by the Ministry of Agriculture Development. Table 5 depicts the changes observed in the yield rate of major crops over the past 50 years.

Table 5. Average yield of crops (kg/ha) in Nepal, 1962 and 2012.

<table>
<thead>
<tr>
<th>Year</th>
<th>Rice</th>
<th>Maize</th>
<th>Wheat</th>
<th>Millet</th>
<th>Mustard seed</th>
<th>Sugarcane</th>
<th>Jute</th>
<th>Potatoes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961/62</td>
<td>2010</td>
<td>1076</td>
<td>946</td>
<td>1312</td>
<td>553</td>
<td>19886</td>
<td>1236</td>
<td>6628</td>
</tr>
<tr>
<td>2011/12</td>
<td>2970</td>
<td>2341</td>
<td>2297</td>
<td>1127</td>
<td>731</td>
<td>39396</td>
<td>n.a.</td>
<td>13041</td>
</tr>
<tr>
<td>Absolute change</td>
<td>960</td>
<td>1265</td>
<td>1351</td>
<td>-185</td>
<td>178</td>
<td>19510</td>
<td>..</td>
<td>6413</td>
</tr>
<tr>
<td>Percent change</td>
<td>47.8</td>
<td>117.6</td>
<td>142.8</td>
<td>-14.1</td>
<td>32.2</td>
<td>98.1</td>
<td>..</td>
<td>96.8</td>
</tr>
</tbody>
</table>

Issues, challenges, lesson learned and the way forward

- Recall periods ranged from one to twelve months from the date of enumeration. There was no mechanism available for enumerators to cross-check estimates with the reported data.
- Production figures reported by farmers were based on their recall or prediction. As such, the figures may not be very accurate.
- Farmers in different parts of the country use different local units of area and weight measurement. The conversion of these figures from local units to standardized units is one of the possible sources of error in the census.
- Farmers often forget to include payments that were made in-kind or were non-monetary (e.g. payments to labourers or gifts to family and friends) in their production estimates.
- A major source of error in long-duration fieldwork and lengthy interviews is that of “enumerator and respondent fatigue”.

127
• Recruitment of field staff was constrained because of government rules that require that 50% of staff be hired on a contract basis, and 50% on secondment of government employees.

• Insufficient transportation expenses and training allowance hindered the census work in some very rural areas of the hills and mountains.

• Respondents were scared to provide real data on the area of agricultural land due to fear of a possible change in the taxation policy of the government that would result in a greater financial burden to them.

• The manual for the community questionnaire, which was introduced in the WPCA 2010, lacks tabulation plans to supplement the information collected at the holding level.

Lessons learned

• The information collected in the Household Listing Form of the 2011 population census was the basis for constructing the PSUs for the agricultural census. However, data from the listing form was not made available on time for the frame construction. It was, therefore, decided to adopt an alternative method. The number of agricultural households was transcribed from the listing form based on “quick counts” of the population census. The number of SSUs was slightly less than the size fixed by the design as a result of this. Such a problem can be overcome if the Population Census Section processes the data on time to meet the agricultural census requirements.

• All enumerators should be recruited afresh rather than be seconded from other government offices.

Way forward

• The CBS has not yet used the latest information technology (IT) for preparing and managing the agricultural census.

• The Crop Production Survey, using the area sampling frame based on remote sensing, will produce more reliable estimates.

• IT tools such as mobile phones and tablets can be used to upload data and serve as storage facilities in real time.

• Access to IT tools and facilities are especially important for gathering and synchronizing data and statistics.
• IT training for the entire census labor force would significantly improve Nepal’s agricultural census.

Conclusion

With the experience and technical expertise gained from the previous five agricultural censuses, Nepal conducted its sixth National Sample Census of Agriculture (2011/12 NSCA) on its own with minimal outside support.

The upcoming agricultural census should be able to respond to the socioeconomic changes that have come about in the country, which is moving towards a federal structure. The quality and availability of data depends upon the capacity of the institutions involved in the generation of statistical information. The CBS still lacks adequate trained labour and, to some extent, funds for undertaking the forthcoming census on its own. Therefore, Nepal may require international cooperation for improving the agriculture census in the areas of technical assistance, training and equipment.

Better coordination with the population census should be established because the frame for the agricultural census is based on the data collected by the population census. Any delay in processing the population census data hampers the schedule of the agricultural census. This aspect needs to be addressed.

The in-house capacity to process census data is still lacking and has to be outsourced. The CAPI (Computer Assisted Personal Interviewing) system should be adopted in the upcoming census. This would certainly expedite the publication of results and reduce errors.

Among others, the agricultural census should provide the framework for drawing the sample needed in implementing the sample survey. This has not been the case in Nepal, however. Current agricultural surveys and other ad hoc surveys are not based on frames provided by the agricultural census.

The present definition of an agricultural holding for the purpose of the agricultural census is the same as the one adopted during the 1991/92 census. There have been substantial changes in the agriculture sector since then, and it would be worthwhile to examine whether the same definition still holds, or needs to be modified for the next census.

Among the three regions — mountains, hills and the tarai — the mountains constitute 35% of the total land area. The mountain region has great potential for livestock raising. A separate livestock census would certainly help in better planning of the agriculture sector of the country.
References


Annex 1

National Sample Census of Agriculture 2011/12 – sample allocation

<table>
<thead>
<tr>
<th>District</th>
<th>DR</th>
<th>Ecological Belt</th>
<th>Agri. Holding</th>
<th>Total VDC/Mun</th>
<th>Selected VDC/Mun</th>
<th>No. of wards</th>
<th>No. of PSUs</th>
<th>Sample PSUs</th>
<th>Sample SSUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Taplejung</td>
<td>EDR</td>
<td>Mountain</td>
<td>26286</td>
<td>50</td>
<td>45</td>
<td>448</td>
<td>396</td>
<td>58</td>
<td>1450</td>
</tr>
<tr>
<td>2 Panchthar</td>
<td>EDR</td>
<td>Hill</td>
<td>41129</td>
<td>41</td>
<td>41</td>
<td>369</td>
<td>367</td>
<td>63</td>
<td>1575</td>
</tr>
<tr>
<td>3 Ilam</td>
<td>EDR</td>
<td>Hill</td>
<td>64296</td>
<td>49</td>
<td>48</td>
<td>440</td>
<td>443</td>
<td>72</td>
<td>1800</td>
</tr>
<tr>
<td>4 Jhapa</td>
<td>EDR</td>
<td>Terai</td>
<td>182036</td>
<td>50</td>
<td>50</td>
<td>470</td>
<td>557</td>
<td>107</td>
<td>2675</td>
</tr>
<tr>
<td>5 Morang</td>
<td>EDR</td>
<td>Terai</td>
<td>213363</td>
<td>66</td>
<td>66</td>
<td>607</td>
<td>733</td>
<td>106</td>
<td>2650</td>
</tr>
<tr>
<td>6 Sunsari</td>
<td>EDR</td>
<td>Terai</td>
<td>161254</td>
<td>52</td>
<td>50</td>
<td>479</td>
<td>584</td>
<td>89</td>
<td>2225</td>
</tr>
<tr>
<td>7 Dhankuta</td>
<td>EDR</td>
<td>Hill</td>
<td>37203</td>
<td>36</td>
<td>36</td>
<td>324</td>
<td>320</td>
<td>67</td>
<td>1675</td>
</tr>
<tr>
<td>8 Terhathum</td>
<td>EDR</td>
<td>Hill</td>
<td>21865</td>
<td>32</td>
<td>32</td>
<td>288</td>
<td>284</td>
<td>58</td>
<td>1450</td>
</tr>
<tr>
<td>9 Sankhuwasabha</td>
<td>EDR</td>
<td>Mountain</td>
<td>34772</td>
<td>34</td>
<td>33</td>
<td>310</td>
<td>301</td>
<td>67</td>
<td>1675</td>
</tr>
<tr>
<td>10 Bhojpur</td>
<td>EDR</td>
<td>Hill</td>
<td>39283</td>
<td>63</td>
<td>53</td>
<td>567</td>
<td>541</td>
<td>68</td>
<td>1700</td>
</tr>
<tr>
<td>11 Solukhumbu</td>
<td>EDR</td>
<td>Mountain</td>
<td>23426</td>
<td>34</td>
<td>32</td>
<td>306</td>
<td>380</td>
<td>56</td>
<td>1400</td>
</tr>
<tr>
<td>12 Okhaldhunga</td>
<td>EDR</td>
<td>Hill</td>
<td>32206</td>
<td>56</td>
<td>54</td>
<td>504</td>
<td>486</td>
<td>59</td>
<td>1475</td>
</tr>
<tr>
<td>13 Khotang</td>
<td>EDR</td>
<td>Hill</td>
<td>42148</td>
<td>76</td>
<td>65</td>
<td>684</td>
<td>644</td>
<td>79</td>
<td>1975</td>
</tr>
<tr>
<td>14 Udayapur</td>
<td>EDR</td>
<td>Hill</td>
<td>64819</td>
<td>45</td>
<td>37</td>
<td>413</td>
<td>407</td>
<td>66</td>
<td>1650</td>
</tr>
<tr>
<td>District</td>
<td>DR</td>
<td>Ecological Belt</td>
<td>Agri. Holding</td>
<td>Total VDC/Mun</td>
<td>Selected VDC/Mun</td>
<td>No.of wards</td>
<td>No. of PSUs</td>
<td>Sample PSUs</td>
<td>Sample SSUs</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Saptari</td>
<td>EDR</td>
<td>Terai</td>
<td>120606</td>
<td>115</td>
<td>86</td>
<td>1035</td>
<td>1038</td>
<td>94</td>
<td>2350</td>
</tr>
<tr>
<td>Siraha</td>
<td>EDR</td>
<td>Terai</td>
<td>119609</td>
<td>108</td>
<td>77</td>
<td>972</td>
<td>957</td>
<td>90</td>
<td>2250</td>
</tr>
<tr>
<td>Dhanusa</td>
<td>CDR</td>
<td>Terai</td>
<td>139342</td>
<td>102</td>
<td>78</td>
<td>925</td>
<td>960</td>
<td>91</td>
<td>2275</td>
</tr>
<tr>
<td>Mahottari</td>
<td>CDR</td>
<td>Terai</td>
<td>111299</td>
<td>77</td>
<td>66</td>
<td>697</td>
<td>695</td>
<td>85</td>
<td>2125</td>
</tr>
<tr>
<td>Sarlahi</td>
<td>CDR</td>
<td>Terai</td>
<td>132297</td>
<td>100</td>
<td>74</td>
<td>901</td>
<td>887</td>
<td>92</td>
<td>2300</td>
</tr>
<tr>
<td>Sindhuli</td>
<td>CDR</td>
<td>Hill</td>
<td>57003</td>
<td>54</td>
<td>46</td>
<td>495</td>
<td>481</td>
<td>68</td>
<td>1700</td>
</tr>
<tr>
<td>Ramechhap</td>
<td>CDR</td>
<td>Hill</td>
<td>43123</td>
<td>55</td>
<td>50</td>
<td>495</td>
<td>491</td>
<td>67</td>
<td>1675</td>
</tr>
<tr>
<td>Dolakha</td>
<td>CDR</td>
<td>Mountain</td>
<td>45402</td>
<td>52</td>
<td>40</td>
<td>472</td>
<td>468</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>Sindhupalchok</td>
<td>CDR</td>
<td>Mountain</td>
<td>65585</td>
<td>79</td>
<td>67</td>
<td>711</td>
<td>690</td>
<td>80</td>
<td>2000</td>
</tr>
<tr>
<td>Kavre</td>
<td>CDR</td>
<td>Hill</td>
<td>80828</td>
<td>90</td>
<td>66</td>
<td>816</td>
<td>786</td>
<td>79</td>
<td>1975</td>
</tr>
<tr>
<td>Lalitpur</td>
<td>CDR</td>
<td>Hill</td>
<td>107073</td>
<td>42</td>
<td>34</td>
<td>391</td>
<td>426</td>
<td>53</td>
<td>1325</td>
</tr>
<tr>
<td>Bhaktapur</td>
<td>CDR</td>
<td>Hill</td>
<td>67727</td>
<td>18</td>
<td>18</td>
<td>178</td>
<td>256</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>Kathmandu</td>
<td>CDR</td>
<td>Hill</td>
<td>415236</td>
<td>59</td>
<td>46</td>
<td>567</td>
<td>667</td>
<td>53</td>
<td>1325</td>
</tr>
<tr>
<td>Nuwakot</td>
<td>CDR</td>
<td>Hill</td>
<td>58900</td>
<td>62</td>
<td>54</td>
<td>560</td>
<td>553</td>
<td>72</td>
<td>1800</td>
</tr>
<tr>
<td>Rasuwa</td>
<td>CDR</td>
<td>Mountain</td>
<td>9718</td>
<td>18</td>
<td>17</td>
<td>162</td>
<td>130</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>Dhading</td>
<td>CDR</td>
<td>Hill</td>
<td>73356</td>
<td>50</td>
<td>48</td>
<td>450</td>
<td>449</td>
<td>71</td>
<td>1775</td>
</tr>
<tr>
<td>Makwanpur</td>
<td>CDR</td>
<td>Hill</td>
<td>85603</td>
<td>44</td>
<td>38</td>
<td>398</td>
<td>457</td>
<td>67</td>
<td>1675</td>
</tr>
<tr>
<td>Rautahat</td>
<td>CDR</td>
<td>Terai</td>
<td>106417</td>
<td>97</td>
<td>70</td>
<td>876</td>
<td>861</td>
<td>89</td>
<td>2225</td>
</tr>
<tr>
<td>Bara</td>
<td>CDR</td>
<td>Terai</td>
<td>108411</td>
<td>99</td>
<td>78</td>
<td>896</td>
<td>888</td>
<td>94</td>
<td>2350</td>
</tr>
<tr>
<td>Parsa</td>
<td>CDR</td>
<td>Terai</td>
<td>96142</td>
<td>83</td>
<td>71</td>
<td>757</td>
<td>796</td>
<td>85</td>
<td>2125</td>
</tr>
<tr>
<td>Chitawan</td>
<td>CDR</td>
<td>Terai</td>
<td>133489</td>
<td>38</td>
<td>36</td>
<td>350</td>
<td>410</td>
<td>83</td>
<td>2075</td>
</tr>
<tr>
<td>District</td>
<td>DR</td>
<td>Ecological Belt</td>
<td>Agri. Holding</td>
<td>Total VDC/Mun</td>
<td>Selected VDC/Mun</td>
<td>No. of wards</td>
<td>No. of PSUs</td>
<td>Sample PSUs</td>
<td>Sample SSUs</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>36 Gorkha</td>
<td>WDR</td>
<td>Hill</td>
<td>65882</td>
<td>67</td>
<td>54</td>
<td>605</td>
<td>573</td>
<td>73</td>
<td>1825</td>
</tr>
<tr>
<td>37 Lamjung</td>
<td>WDR</td>
<td>Hill</td>
<td>41483</td>
<td>61</td>
<td>53</td>
<td>549</td>
<td>508</td>
<td>68</td>
<td>1700</td>
</tr>
<tr>
<td>38 Tanahu</td>
<td>WDR</td>
<td>Hill</td>
<td>77702</td>
<td>47</td>
<td>46</td>
<td>425</td>
<td>436</td>
<td>72</td>
<td>1800</td>
</tr>
<tr>
<td>39 Syangja</td>
<td>WDR</td>
<td>Hill</td>
<td>68974</td>
<td>62</td>
<td>58</td>
<td>564</td>
<td>585</td>
<td>83</td>
<td>2075</td>
</tr>
<tr>
<td>40 Kaski</td>
<td>WDR</td>
<td>Hill</td>
<td>124836</td>
<td>45</td>
<td>41</td>
<td>420</td>
<td>561</td>
<td>77</td>
<td>1925</td>
</tr>
<tr>
<td>41 Manang</td>
<td>WDR</td>
<td>Mountain</td>
<td>1389</td>
<td></td>
<td></td>
<td>117</td>
<td>50</td>
<td>1250</td>
<td></td>
</tr>
<tr>
<td>42 Mustang</td>
<td>WDR</td>
<td>Mountain</td>
<td>3306</td>
<td>16</td>
<td>16</td>
<td>144</td>
<td>69</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>43 Myagdi</td>
<td>WDR</td>
<td>Hill</td>
<td>27484</td>
<td>41</td>
<td>36</td>
<td>369</td>
<td>334</td>
<td>52</td>
<td>1300</td>
</tr>
<tr>
<td>44 Parbat</td>
<td>WDR</td>
<td>Hill</td>
<td>35526</td>
<td>55</td>
<td>51</td>
<td>495</td>
<td>466</td>
<td>62</td>
<td>1550</td>
</tr>
<tr>
<td>45 Baglung</td>
<td>WDR</td>
<td>Hill</td>
<td>61055</td>
<td>60</td>
<td>53</td>
<td>542</td>
<td>534</td>
<td>73</td>
<td>1825</td>
</tr>
<tr>
<td>46 Gulmi</td>
<td>WDR</td>
<td>Hill</td>
<td>64518</td>
<td>79</td>
<td>62</td>
<td>711</td>
<td>700</td>
<td>65</td>
<td>1625</td>
</tr>
<tr>
<td>47 Palpa</td>
<td>WDR</td>
<td>Hill</td>
<td>59630</td>
<td>66</td>
<td>54</td>
<td>600</td>
<td>582</td>
<td>64</td>
<td>1600</td>
</tr>
<tr>
<td>48 Nawalparasi</td>
<td>WDR</td>
<td>Terai</td>
<td>128480</td>
<td>74</td>
<td>60</td>
<td>670</td>
<td>664</td>
<td>89</td>
<td>2225</td>
</tr>
<tr>
<td>49 Rupandehi</td>
<td>WDR</td>
<td>Terai</td>
<td>165911</td>
<td>71</td>
<td>63</td>
<td>649</td>
<td>698</td>
<td>99</td>
<td>2475</td>
</tr>
<tr>
<td>50 Kapilbastu</td>
<td>WDR</td>
<td>Terai</td>
<td>90988</td>
<td>78</td>
<td>67</td>
<td>707</td>
<td>704</td>
<td>96</td>
<td>2400</td>
</tr>
<tr>
<td>51 Arghakhanchi</td>
<td>WDR</td>
<td>Hill</td>
<td>46475</td>
<td>42</td>
<td>42</td>
<td>378</td>
<td>378</td>
<td>61</td>
<td>1525</td>
</tr>
<tr>
<td>52 Pyuthan</td>
<td>MWDR</td>
<td>Hill</td>
<td>47530</td>
<td>49</td>
<td>47</td>
<td>441</td>
<td>441</td>
<td>58</td>
<td>1450</td>
</tr>
<tr>
<td>53 Rolpa</td>
<td>MWDR</td>
<td>Hill</td>
<td>43142</td>
<td>51</td>
<td>48</td>
<td>459</td>
<td>454</td>
<td>56</td>
<td>1400</td>
</tr>
<tr>
<td>54 Rukum</td>
<td>MWDR</td>
<td>Hill</td>
<td>41285</td>
<td>43</td>
<td>39</td>
<td>387</td>
<td>384</td>
<td>63</td>
<td>1575</td>
</tr>
<tr>
<td>55 Salyan</td>
<td>MWDR</td>
<td>Hill</td>
<td>46211</td>
<td>47</td>
<td>46</td>
<td>423</td>
<td>418</td>
<td>67</td>
<td>1675</td>
</tr>
<tr>
<td>56 Dang</td>
<td>MWDR</td>
<td>Terai</td>
<td>119120</td>
<td>41</td>
<td>40</td>
<td>373</td>
<td>413</td>
<td>91</td>
<td>2275</td>
</tr>
<tr>
<td>District</td>
<td>DR</td>
<td>Ecological Belt</td>
<td>Agri. Holding</td>
<td>Total VDC/Mun</td>
<td>Selected VDC/Mun</td>
<td>No.of wards</td>
<td>No. of PSUs</td>
<td>Sample PSUs</td>
<td>Sample SSUs</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>57 Banke</td>
<td>MWDR</td>
<td>Terai</td>
<td>94356</td>
<td>47</td>
<td>44</td>
<td>430</td>
<td>441</td>
<td>81</td>
<td>2025</td>
</tr>
<tr>
<td>58 Bardiya</td>
<td>MWDR</td>
<td>Terai</td>
<td>83179</td>
<td>32</td>
<td>32</td>
<td>293</td>
<td>312</td>
<td>84</td>
<td>2100</td>
</tr>
<tr>
<td>59 Surkhet</td>
<td>MWDR</td>
<td>Hill</td>
<td>72747</td>
<td>51</td>
<td>44</td>
<td>462</td>
<td>499</td>
<td>72</td>
<td>1800</td>
</tr>
<tr>
<td>60 Dāilekh</td>
<td>MWDR</td>
<td>Hill</td>
<td>48837</td>
<td>56</td>
<td>49</td>
<td>504</td>
<td>499</td>
<td>63</td>
<td>1575</td>
</tr>
<tr>
<td>61 Jajarkot</td>
<td>MWDR</td>
<td>Hill</td>
<td>30289</td>
<td>30</td>
<td>30</td>
<td>270</td>
<td>269</td>
<td>53</td>
<td>1325</td>
</tr>
<tr>
<td>62 Dolpa</td>
<td>MWDR</td>
<td>Mountain</td>
<td>7313</td>
<td>23</td>
<td>22</td>
<td>207</td>
<td>160</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>63 Jumla</td>
<td>MWDR</td>
<td>Mountain</td>
<td>19404</td>
<td>30</td>
<td>30</td>
<td>270</td>
<td>261</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>64 Kalikot</td>
<td>MWDR</td>
<td>Mountain</td>
<td>22928</td>
<td>30</td>
<td>29</td>
<td>270</td>
<td>266</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>65 Mugu</td>
<td>MWDR</td>
<td>Mountain</td>
<td>9476</td>
<td>24</td>
<td>23</td>
<td>216</td>
<td>174</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>66 Humla</td>
<td>MWDR</td>
<td>Mountain</td>
<td>9383</td>
<td>27</td>
<td>27</td>
<td>243</td>
<td>183</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>67 Bajura</td>
<td>FWDR</td>
<td>Mountain</td>
<td>24433</td>
<td>27</td>
<td>27</td>
<td>243</td>
<td>240</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>68 Bajhang</td>
<td>FWDR</td>
<td>Mountain</td>
<td>33447</td>
<td>47</td>
<td>44</td>
<td>423</td>
<td>413</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>69 Achham</td>
<td>FWDR</td>
<td>Hill</td>
<td>47855</td>
<td>75</td>
<td>52</td>
<td>675</td>
<td>637</td>
<td>52</td>
<td>1300</td>
</tr>
<tr>
<td>70 Doti</td>
<td>FWDR</td>
<td>Hill</td>
<td>41055</td>
<td>51</td>
<td>42</td>
<td>463</td>
<td>440</td>
<td>56</td>
<td>1400</td>
</tr>
<tr>
<td>71 Kailali</td>
<td>FWDR</td>
<td>Terai</td>
<td>140772</td>
<td>44</td>
<td>42</td>
<td>401</td>
<td>466</td>
<td>103</td>
<td>2575</td>
</tr>
<tr>
<td>72 Kanchanpur</td>
<td>FWDR</td>
<td>Terai</td>
<td>81657</td>
<td>20</td>
<td>20</td>
<td>180</td>
<td>214</td>
<td>88</td>
<td>2200</td>
</tr>
<tr>
<td>73 Dadeldhura</td>
<td>FWDR</td>
<td>Hill</td>
<td>26516</td>
<td>21</td>
<td>21</td>
<td>191</td>
<td>192</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>74 Baitadi</td>
<td>FWDR</td>
<td>Hill</td>
<td>44888</td>
<td>63</td>
<td>46</td>
<td>571</td>
<td>554</td>
<td>51</td>
<td>1275</td>
</tr>
<tr>
<td>75 Darchula</td>
<td>FWDR</td>
<td>Mountain</td>
<td>24557</td>
<td>41</td>
<td>38</td>
<td>369</td>
<td>346</td>
<td>50</td>
<td>1250</td>
</tr>
<tr>
<td>Nepal</td>
<td></td>
<td></td>
<td>5385271</td>
<td>3960</td>
<td>3440</td>
<td>36022</td>
<td>36257</td>
<td>5200</td>
<td>130000</td>
</tr>
<tr>
<td>District</td>
<td>DR</td>
<td>Ecological Belt</td>
<td>Agri. Holding</td>
<td>Total VDC/Mun</td>
<td>Selected VDC/Mun</td>
<td>No. of wards</td>
<td>No. of PSUs</td>
<td>Sample PSUs</td>
<td>Sample SSUs</td>
</tr>
<tr>
<td>----------</td>
<td>----</td>
<td>-----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>EDR</td>
<td></td>
<td>Eastern Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDR</td>
<td></td>
<td>Eastern Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDR</td>
<td></td>
<td>Central Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WDR</td>
<td></td>
<td>Western Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MWDR</td>
<td></td>
<td>Mid-Western Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FWDR</td>
<td></td>
<td>Far-Western Development Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Manang District was completely enumerated and the number of selected PSUs was around 50.
Annex 2

Sample design and estimation specifications

Sample design and selection\textsuperscript{23}

The sample design of the 2011/12 agricultural census is a stratified two-stage sampling with districts as strata, wards as first-stage units (FSUs) in rural areas, and enumeration areas (EAs) as FSUs in urban areas. In the following, FSUs in rural and urban areas are collectively called “EAs”. Agricultural holdings are considered as second-stage units (SSUs).

Within a specific district \( k \), the sample is selected as follows:

First-stage units: \( n_k \) EAs are selected with probability \( P_k \), proportional to size (the estimated number of holdings \( M_{kij} \)) systematic (PPS systematic) sampling;

Second-stage units: within each selected EA, \( m_{kij} \) holdings are selected with equal probability systematic sampling and implicit stratification.

The total sample size of EAs across the country was estimated to be 5,200, with 25 SSUs within each EA (a target number of sample holdings). District were taken as strata and 5,200 were allocated to different strata, following compromise power allocation (in proportion to \( x^2 \), where \( x \) is the size measure and the parameter \( \lambda = 0.4 \)).

The holding numbers in each EA were estimated from the 2011 population census. As in the 1992 agricultural census, the sample was designed to be self-weighting within each district; that is, all holdings within a district have the same chance of selection in the sample. The procedure for achieving this is available in a technical report on “Reliability of Data” published by Nepal’s Central Bureau of Statistics in 1994.\textsuperscript{24}


In order to achieve the self-weighting process, the number of holdings, $m_{ki}$, was designed to be fixed as:

$$m_{ki} = 25 \frac{M_{ki}}{M_{ki}'}$$

where,

- $M_{ki}'$ = number of agricultural households in EA i in district k as estimated from the 2011 population census,
- $M_{ki}$ = actual number of holdings in EA i in district k as recorded in the 2012 agricultural census enumeration.

It was expected that, $M_{ki}$ and $M_{ki}'$ would usually be almost the same. For operational purposes, the number of selected holdings was set at 25 if:

$$0.98 < \frac{M_{ki}'}{M_{ki}} < 1.02$$

If this relationship was not fulfilled, then field staff were required to calculate the required number of holdings, $m_{ki}$, according to the above expression, using procedures given in instruction manuals.

The PPS selection of the sample of EAs within each district was undertaken as follows:

Define,

$N$ = number of EAs in the district,

$n$ = number of EAs to be selected in the district,

$z_i$ = the measure of size (MoS: number of agricultural households in this case) for the $i^{th}$ EA in the district,

$$Z = \sum_{i=1}^{N} z_i,$$

$$p_i = \frac{z_i}{Z}, \quad i = 1, 2, 3, \ldots, N,$$

---

\[
\pi_i = \frac{1}{p_i}, \quad i = 1, 2, 3, \ldots \ldots, N.
\]

The \( \pi_i \) values are the selection probabilities for the \( i \)th EA. The PPS systematic sampling selection procedure is described in the following steps.

**Step 1:** In this step, the procedure of implicit stratification is followed. For this purpose, administrative posts (PAs) are considered as implicit strata. Sort list of EAs in the stratum by PAs. Within a PA, arrange the EAs in ascending order of MoS; then in the next PA arrange the EAs in descending order of MoS. Continue this sorting by alternating between ascending and descending order sorting from one PA to the next. This type of sorting helps in improving the efficiency of PPS systematic sampling.

**Step 2:** Check that \( np_i < 1 \), i.e. \( z_i \) is less than \( \frac{Z}{n} \) for all \( i \) in the stratum.

**Step 3:** Compute cumulative total

\[ C_1 = \pi_1, \]
\[ C_2 = C_1 + \pi_1, \]

\[ \ldots \ldots \]

\[ C_{N-1} = C_{N-2} + \pi_{N-1}, \]

\[ C_N = C_{N-1} + \pi_N \] (Note that \( C_N = n \)).

**Step 4:** Generate a random number “r” between 0 and 1. Compute the numbers \( r_i = r+i-1, \) with \( i = 1, 2, 3, \ldots, n+2. \)
Step 5: Select the n EAs with the labels $i_1, i_2, i_3, \ldots, i_n$ such that:

\[ C_{i_{i-1}} < r_i \leq C_{i_i}, \]

\[ C_{i_{i-1}} < r_2 \leq C_{i_2}, \]

\[ C_{i_{i-1}} < r_3 \leq C_{i_3}, \]

\[ \ldots \quad \ldots \]

\[ C_{i_{i-1}} < r_n \leq C_{i_n}. \]

The procedure yields a sample of size n with PPS systematic sampling and the selection probabilities are given by $\pi_i = np_i$; $i=1, 2, 3, \ldots, N$.

For each selected EA, a list of holdings is prepared and ordered according by stratum (see details in Chapter 1). A systematic random sample is selected by applying a sampling interval of $I$ to the holdings in the EA, where $I$ is calculated as:

\[ I = \frac{\text{number of holdings recorded in the EA in the 2012 census listing}}{\text{number of holdings to be selected in the EA}} \]

**Estimation**

All parameters are estimated at the district level first; development region, ecological belt and national estimates are obtained by aggregating across districts.
Estimation of totals

Sample weights – Sampling weights are needed for developing the estimates for various parameters (such as population total or mean). Estimates for such parameters are linear in nature with sample observations suitably weighted with appropriate sampling weights. The unit of observation in the 2012 census is operational holding, whereas sampling units are agricultural households. The weighting procedure is essentially based on three types of weights: base weights, non-response adjustments, and post-stratification adjustments.

Base weights – Base weights are the inverse of selection probabilities for individual holdings, which are the units of observation. Because the selection probabilities are associated with the units of selection, which in this case are agricultural households, the agricultural households associated with the holding provide the base weights for the holdings. In Nepal’s context, “the holding is generally the same as a household”.

In two-stage sampling, the selection probability of an SSU is the product of selection probability of a corresponding PSU and the conditional selection probability of an SSU for the given PSU. In the present case, EAs are PSUs that are selected with PPS systematic sampling, and agricultural households are SSUs that are selected with equal probability sampling.

Let $\pi_i$ be the probability of selection for $i^{th}$ PSU (i.e. EA) and $\pi_{ij}$ be the conditional probability for selecting $j^{th}$ SSU (household) in the $i^{th}$ PSU, then the probability of selection for $j^{th}$ SSU in $i^{th}$ PSU is given by

$$\pi_{ij} = \pi_i \pi_{ji}.$$ 

In this case, $\pi_i = n \frac{X_i}{X}$, where,

$X_i$ is the measure of size (number of agricultural holdings in $i^{th}$ EA as per the 2011 population census), and

---

X is the sum of $X_i$ in the specific stratum to which $i^{th}$ EA belongs.

Also, $\pi_{ij} = \frac{m}{M_i}$,

where,

$M_i$ is the number of agricultural holdings in $i^{th}$ EA as observed at the time of field work for preparing the frame, and

$m$ is the number of households selected in each EA.

Thus, $\pi_j = \frac{nmX_i}{XM_i}$

in the case when $X_i = M_i$, $\pi_j = \frac{nm}{X}$.

In general, the base weights for each household in the $i^{th}$ EA is $\frac{XM_i}{nmX_i}$.

**Non-response adjustments** – The sampling weights described above are based on the planned sample sizes. However, invariably, there is some amount of non-response in every survey, which disturbs the weights. Therefore, there is a need for adjusting for non-response. Normally, non-response adjustments are done within each EA. The adjustment factor is $(m/r)$, where $m$ is the number of sampled holdings while $r$ is the number of responding households.

**Final weights** – The final weights are the product of base weight and non-response adjustments.

**Estimation of error**

The sampling designed followed in the 2012 agricultural census is broadly similar to that of the 1992 agricultural census, and the estimation procedure for sampling variances is similar to that described in the report “Reliability of Data” published by the CBS in 1994. The following is the procedure described in this report.

---

Standard errors were estimated using the subsample method. In each district, sample EAs were assigned to 10 subsamples, with the same number of EAs in each subsample.

To estimate the standard errors on the estimate of average per holding for characteristics $X$ in district $k$, the estimate of average is first calculated for each subsample $g$ as follows;

$$
\bar{x}_{kg} = \frac{n_g/10 \sum_{j=1}^{n_g} X_{kij}}{\sum_{j=1}^{n_g} m_{ki}}
$$

where:

$x_{kij}^{(g)}$ = the value of characteristics $x$ in district $k$, EA $i$ and holding $j$ for subsample $g$ ($g = 1, 2, \ldots, 10$);

The standard error of the estimate of average per holding for characteristics $X$ in district $k$ is given by;

$$
s(\bar{x}_k) = \sqrt{\frac{\sum_{g=1}^{10} (\bar{x}_{kg} - \bar{x}_k)^2}{90}},
$$

where $\bar{x}_k = \frac{\sum_{g=1}^{10} \bar{x}_{kg}}{10}$

The standard errors of the estimate of total for characteristic $X$ in district $k$ is given by;

$$
s(x_k) = M_k s(\bar{x}_k)
$$

The standard error of the estimate of total for characteristic $X$ at the national level is given by;

$$
s(x') = \sqrt{\sum_{k=1}^{75} s^2(x_k)}$$
The standard error of the estimate of average per holding for characteristic $X$ at
the national level is given by:

$$ s(\bar{x}) = \frac{s(x')}{\sqrt{\sum_{k=1}^{n_k} M'_k}} $$

Standard errors for ecological belts and development regions are formed
aggregating across the relevant districts making up the area in the same way as
for standard errors on national estimates.

**Estimation of sample design parameters**

The estimation procedures of sample design parameters is similar to that of the
1992 agricultural census described in the “Reliability of Data” by the CBS

The design effect measures the variance of an estimate in comparison with the
variance which would have been obtained if simple random sampling had been
used. The design effect $d_k$ for characteristics $X$ in district $k$ is estimated as:

$$ d_k = \frac{s^2(\bar{x}_k) x m_k}{s^2(x_{kij})} $$

where:

$$ s^2(x_{kij}) = \frac{1}{m_k - 1} \sum_{j=1}^{n_k} \sum_{i=1}^{m_{ki}} (x_{kij} - \bar{x}_k)^2 $$

$$ m_k = \sum_{j=1}^{n_k} m_{ki} $$

The coefficient of variation is given by:

$$ cv_k = \frac{s(x_{kij})}{\bar{x}_k} $$

The measure of homogeneity for characteristic $X$ in district $k$ is estimated as:

$$ \delta_k = \frac{d_k - 1}{\bar{m}_k - 1} $$
where:

\[ \bar{m}_k = \text{average sample holdings per EA in district } k. \]

\( \delta_k \) is a measure of the relationship between the variability of the first and second stages of sampling. If the variability within EAs is high in comparison with the variability between EAs, then \( \delta_k \) will be small. If, on the other hand, EAs are very homogeneous, then \( \delta_k \) will be high.

\( \delta_k \) is influenced by the size of EAs — the larger the EAs are, the more heterogeneous will they be and, therefore, \( \delta_k \) will be lower.

In assessing the sample design for future censuses, decisions will need to take on how many EAs to sample, and then how many holdings to sample within each selected EA. This decision is based on variance and cost (or time) factors.

The total cost of conducting the census enumeration in district \( k \) can be represented as:

\[ C_k = C_{ko} + n_k C_{k1} + n_k \bar{m}_k C_{k2} \]

where:

- \( C_{ko} \) = overhead costs
- \( C_{k1} \) = average costs associated with each of the first stage units (e.g. listing, travel to EAs); and
- \( C_{k2} \) = average costs associated with each of the second stage units in each EA (e.g. interviewing holdings).

The optimum number of sample holdings to sample per EA is calculated as:

\[ \text{opt} (\bar{m}_k) = \frac{C_{k1}}{C_{k2} \delta_k} \left( 1 - \frac{\delta_k}{\bar{m}_k} \right) \]

A low \( C_{k2} \) means interviewer costs are low and, therefore, \( \bar{m}_k \) should be high.

A high \( C_{k2} \) means that interviewer costs are high and, therefore, \( \bar{m}_k \) should be
low. High within EA variability means a low $\delta_k$, implying the need for a large $\overline{m}_k$. Low within EA variability means a high $\delta_k$ and, therefore, a low $\overline{m}_k$.

**Estimation procedure**

The estimation procedure is given in Box 6.5. All parameters were estimated at the district level. The higher levels (i.e. zonal, regional and national estimates) were obtained by summing up through district estimates or weighting them respectively.

If,

$X_k = \text{estimate of the average value of characteristic } X \text{ per holding},$

$X_k = \text{estimate of the total value of characteristic } X,$

$r_k = \text{estimate of ratio of characteristics } X \text{ and } Y,$

Then, the estimates of the parameters at national level were obtained as:

$$
\bar{X} = \frac{\sum_{k=1}^{75} X_k}{\sum_{k=1}^{75} M_k}; \quad X = \sum_{k=1}^{75} X_k; \quad R = \frac{X}{Y}.
$$

For sampling error estimation, estimates for subsamples were used. Thus, for each district ten subsample estimates were obtained and the standard error for the estimate $X$ was calculated as:

$$
s(\bar{x}_d) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{10} (\bar{x}_{di} - \bar{x}_d)^2};
$$

where, $n = 10$ and $\bar{x}_d = \frac{\sum_{i=1}^{10} x_{di}}{n}$.
National level estimates

Estimate of the average value of characteristic $X$ per holding in district $k$ is given by:

$$\bar{x} = \frac{\sum_{k=1}^{75} M_k' \bar{x}_k}{\sum_{k=1}^{75} M_k'}$$

where,

$M_k' =$ estimated number of holdings (from population census) in district $k$,

$\bar{x}_k =$ estimated average value of characteristic $X$ per holding in district $k$.

The estimate of the total value of characteristic $X$ in district $k$ is given by:

$$x' = \sum_{k=1}^{75} M_k' \bar{x}_k$$

The estimate of the ratio of characteristics $X$ and $Y$ in district $k$ is given by:

$$r = \frac{x'}{y'}$$

Estimate of development regions and ecological belts are formed by aggregating across the relevant districts making up the area in the same way as for national estimates.
Estimation of Standard Errors

Standard errors were estimated using the sub-sample method. In each district, sample EAs were assigned to ten subsamples, with the same number of EAs in each subsample.

To estimate the standard errors on the estimate of average per holding for characteristics $X$ in district $k$, the estimate of average is first calculated for each subsample $g$ as follows:

$$
\bar{x}_{kg} = \frac{\sum_{i=1}^{n_k/10} \sum_{j=1}^{m_{ki}} x^{(g)}_{kj}}{\sum_{j=1}^{m_{ki}} m_{ki}}
$$

where:

$x^{(g)}_{kj} = \text{the value of characteristics } x \text{ in district } k, \text{ EA } i \text{ and holding } j \text{ for subsample } g \text{ (g = 1, 2, \ldots, 10)}$;

The standard error of the estimate of average per holding for characteristics $X$ in district $k$ is given by:

$$
s(\bar{x}_k) = \sqrt{\frac{\sum_{g=1}^{10} (\bar{x}_{kg} - \bar{x}_k)^2}{90}},
$$

where $\bar{x}_k = \frac{\sum_{g=1}^{10} \bar{x}_{kg}}{10}$

The standard error of the estimate of total for characteristic $X$ in district $k$ is given by:

$$
s(x'_k) = M' s(\bar{x}_k)
$$
The standard error of the estimate of total for characteristic \( X \) at the national level is given by:

\[
s(x') = \sqrt{\sum_{k=1}^{75} s^2(x'_k)}
\]

The standard error of the estimate of average per holding for characteristic \( X \) at the national level is given by:

\[
s(\bar{x}) = \frac{s(x')}{\sqrt{\sum_{k=1}^{75} M'_k}}
\]

Standard errors for ecological belts and development regions are formed aggregating across the relevant districts making up the area in the same way as for standard errors on national estimates.

**Estimation of sample design parameters**

The design effect measures the variance of an estimate in comparison with the variance that would have been obtained if simple random sampling had been used. The design effect \( d_k \) for characteristics \( X \) in district \( k \) is estimated as:

\[
d_k = \frac{s^2(\bar{x}_k)xm_k}{s^2(x_{ki})}
\]

where:

\[
s^2(x_{ki}) = \frac{1}{m_k - 1} \sum_{j=1}^{n_k} \sum_{j'=1}^{m_k} (x_{kij} - \bar{x}_k)^2
\]

\[
m_k = \sum_{j=1}^{n_k} m_{ki}
\]

The coefficient of variation is given by:

\[
cv_k = \frac{s(x_{ki})}{\bar{x}_k}
\]
The measure of homogeneity for characteristic $X$ in district $k$ is estimated as:

$$
\delta_k = \frac{d_k - 1}{m_k - 1}
$$

where:

$m_k$ = average sample holdings per EA in district $k$.

$\delta_k$ is a measure of the relationship between the variability of the first and second stages of sampling. If the variability within EAs is high in comparison with the variability between EAs, then $\delta_k$ will be small. If, on the other hand, EAs are very homogeneous, then $\delta_k$ will be high.

$\delta_k$ is influenced by the size of EAs; that is, the larger that EAs are, the more heterogeneous will they be and, therefore, $\delta_k$ will be lower.
## Annex 3

National Sample Census of Agriculture, 2011/12, standard errors – ecological belts

<table>
<thead>
<tr>
<th>Data item</th>
<th>Census estimate</th>
<th>Standard error</th>
<th>Relative standard error</th>
<th>Census estimate</th>
<th>Standard error</th>
<th>Relative standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total area of holdings (ha)</td>
<td>213931.5</td>
<td>3624</td>
<td>1.7</td>
<td>986073.2</td>
<td>8061</td>
<td>0.8</td>
</tr>
<tr>
<td>Area of temporary crops (ha)</td>
<td>170375</td>
<td>5110</td>
<td>1.7</td>
<td>733430</td>
<td>10714</td>
<td>0.8</td>
</tr>
<tr>
<td>Area under paddy (ha)</td>
<td>61,725</td>
<td>1643</td>
<td>2.7</td>
<td>309,454</td>
<td>2999</td>
<td>1.0</td>
</tr>
<tr>
<td>Area under maize (ha)</td>
<td>74,186</td>
<td>1443</td>
<td>9.3</td>
<td>453,129</td>
<td>4375</td>
<td>1.0</td>
</tr>
<tr>
<td>Area under wheat (ha)</td>
<td>56,731</td>
<td>1655</td>
<td>2.9</td>
<td>185,887</td>
<td>2142</td>
<td>1.2</td>
</tr>
<tr>
<td>Area under potato (ha)</td>
<td>18,620</td>
<td>955</td>
<td>5.1</td>
<td>48,747</td>
<td>1634</td>
<td>3.4</td>
</tr>
<tr>
<td>Area of irrigated land (ha)</td>
<td>58378</td>
<td>1718</td>
<td>2.9</td>
<td>270273</td>
<td>2645</td>
<td>1.0</td>
</tr>
<tr>
<td>Total paddy production (qtl)</td>
<td>1,308,629</td>
<td>47149</td>
<td>3.6</td>
<td>8,192,413</td>
<td>100364</td>
<td>1.2</td>
</tr>
<tr>
<td>Total maize production (qtl)</td>
<td>1,589,168</td>
<td>32805</td>
<td>2.1</td>
<td>10,189,152</td>
<td>98085</td>
<td>1.0</td>
</tr>
<tr>
<td>Total wheat production (qtl)</td>
<td>951,121</td>
<td>27773</td>
<td>2.9</td>
<td>3,721,162</td>
<td>41411</td>
<td>1.1</td>
</tr>
<tr>
<td>Total potato production (qtl)</td>
<td>2,293,906</td>
<td>122750</td>
<td>5.4</td>
<td>6,359,622</td>
<td>211633</td>
<td>3.3</td>
</tr>
<tr>
<td>Farm population</td>
<td>1695538</td>
<td>10528</td>
<td>0.6</td>
<td>8614984</td>
<td>25305</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Tarai</td>
<td>Total Nepal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total area of holdings (ha)</td>
<td>1325634.5</td>
<td>2,525,639</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of temporary crops (ha)</td>
<td>1219492</td>
<td>2,123,297</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under paddy (ha)</td>
<td>1,084,803</td>
<td>1,455,983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under maize (ha)</td>
<td>146,389</td>
<td>673,704</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under wheat (ha)</td>
<td>506,777</td>
<td>749,395</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under potato (ha)</td>
<td>38,215</td>
<td>105,582</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of irrigated land (ha)</td>
<td>984756</td>
<td>1,313,406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total paddy production (qtl)</td>
<td>33,737,750</td>
<td>4,323,879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total maize production (qtl)</td>
<td>3,993,171</td>
<td>1,577,149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total wheat production (qtl)</td>
<td>12,542,530</td>
<td>1,721,481</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total potato production (qtl)</td>
<td>5,115,102</td>
<td>1,376,863</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farm population</td>
<td>10242020</td>
<td>20,552,543</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annex 4

National Sample Census of Agriculture 2011/12, standard errors by development region

<table>
<thead>
<tr>
<th></th>
<th>Census estimate</th>
<th>Standard error</th>
<th>Relative standard error %</th>
<th>Census estimate</th>
<th>Standard error</th>
<th>Relative standard error %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eastern</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>Central</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total area of holdings (ha)</td>
<td>755,178</td>
<td>10,460</td>
<td>1.4</td>
<td>716,861</td>
<td>7,756</td>
<td>1.1</td>
</tr>
<tr>
<td>Area of temporary crops</td>
<td>628,976</td>
<td>16,612</td>
<td>1.5</td>
<td>637,970</td>
<td>13,554</td>
<td>1.1</td>
</tr>
<tr>
<td>Area under paddy (ha)</td>
<td>482,006</td>
<td>6,661</td>
<td>1.4</td>
<td>447,348</td>
<td>5,683</td>
<td>1.3</td>
</tr>
<tr>
<td>Area under maize (ha)</td>
<td>190,458</td>
<td>2,638</td>
<td>1.4</td>
<td>206,551</td>
<td>3,409</td>
<td>1.7</td>
</tr>
<tr>
<td>Area under wheat (ha)</td>
<td>157,585</td>
<td>4,227</td>
<td>2.7</td>
<td>229,266</td>
<td>3,518</td>
<td>1.5</td>
</tr>
<tr>
<td>Area under potato (ha)</td>
<td>37,467</td>
<td>1,535</td>
<td>4.1</td>
<td>35,791</td>
<td>1,399</td>
<td>3.9</td>
</tr>
<tr>
<td>Area of irrigated land (ha)</td>
<td>394,392</td>
<td>7,487</td>
<td>1.9</td>
<td>429,457</td>
<td>5,724</td>
<td>1.3</td>
</tr>
<tr>
<td>Total paddy production (qtl)</td>
<td>13,030,030</td>
<td>232,813</td>
<td>1.8</td>
<td>14,530,407</td>
<td>233,239</td>
<td>1.6</td>
</tr>
<tr>
<td>Total maize production (qtl)</td>
<td>4,245,475</td>
<td>66,905</td>
<td>1.6</td>
<td>5,271,785</td>
<td>101,060</td>
<td>1.9</td>
</tr>
<tr>
<td>Total wheat production (qtl)</td>
<td>3,396,204</td>
<td>101,554</td>
<td>3.0</td>
<td>5,676,893</td>
<td>119,292</td>
<td>2.1</td>
</tr>
<tr>
<td>Total potato production (qtl)</td>
<td>4,881,499</td>
<td>200,065</td>
<td>4.1</td>
<td>4,836,317</td>
<td>178,672</td>
<td>3.7</td>
</tr>
<tr>
<td>Farm population</td>
<td>4,634,993</td>
<td>17,879</td>
<td>0.4</td>
<td>6,366,835</td>
<td>28,188</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Western</td>
<td>Mid-western</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total area of holdings (ha)</td>
<td>482,548</td>
<td>353,624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of temporary crops (ha)</td>
<td>371,738</td>
<td>298,960</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under paddy (ha)</td>
<td>244,834</td>
<td>153,011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under maize (ha)</td>
<td>136,411</td>
<td>108,902</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under wheat (ha)</td>
<td>97,722</td>
<td>139,594</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under potato (ha)</td>
<td>14,159</td>
<td>11,176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of irrigated land (ha)</td>
<td>209,769</td>
<td>152,516</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total paddy production (qtl)</td>
<td>7,417,847</td>
<td>4,651,449</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total maize production (qtl)</td>
<td>3,345,199</td>
<td>2,272,294</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total wheat production (qtl)</td>
<td>2,544,386</td>
<td>3,056,249</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total potato production (qtl)</td>
<td>1,807,122</td>
<td>1,415,417</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farm population</td>
<td>3,999,572</td>
<td>3,158,172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far-western</td>
<td>Total Nepal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total area of holdings (ha)</td>
<td>217,430</td>
<td>2,525,639</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of temporary crops (ha)</td>
<td>185,653</td>
<td>2,123,297</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under paddy (ha)</td>
<td>128,784</td>
<td>1,455,983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under maize (ha)</td>
<td>31,382</td>
<td>673,704</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under wheat (ha)</td>
<td>125,227</td>
<td>749,395</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area under potato (ha)</td>
<td>6,989</td>
<td>105,582</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of irrigated land (ha)</td>
<td>127,272</td>
<td>1,313,406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total paddy production (qtl)</td>
<td>3,609,059</td>
<td>4,323,879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total maize production (qtl)</td>
<td>636,738</td>
<td>1,577,149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total wheat production (qtl)</td>
<td>2,541,081</td>
<td>1,721,481</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total potato production (qtl)</td>
<td>828,275</td>
<td>1,376,863</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farm population</td>
<td>2,392,971</td>
<td>20,552,543</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4,388</td>
<td>16,044</td>
</tr>
<tr>
<td></td>
<td>7,780</td>
<td>26,243</td>
</tr>
<tr>
<td></td>
<td>2,816</td>
<td>10,734</td>
</tr>
<tr>
<td></td>
<td>1,119</td>
<td>5,308</td>
</tr>
<tr>
<td></td>
<td>3,012</td>
<td>7,164</td>
</tr>
<tr>
<td></td>
<td>226</td>
<td>2,180</td>
</tr>
<tr>
<td></td>
<td>3,173</td>
<td>11,045</td>
</tr>
<tr>
<td></td>
<td>72,632</td>
<td>390,662</td>
</tr>
<tr>
<td></td>
<td>22,629</td>
<td>139,531</td>
</tr>
<tr>
<td></td>
<td>68,062</td>
<td>198,239</td>
</tr>
<tr>
<td></td>
<td>24,678</td>
<td>281,142</td>
</tr>
<tr>
<td></td>
<td>15,339</td>
<td>46,183</td>
</tr>
</tbody>
</table>
Abstract

Based on the achievements and experience of the National Institute of Statistics of Rwanda in conducting agricultural surveys, this paper presents the agricultural survey methodology developed after taking into account the availability and sustainability of resources.

The paper presents a methodology that moved beyond traditional agricultural sampling to address not only the local context and to meet the Global Strategies for Improving Agricultural and Rural Statistics, but also to provide data-users and decision-makers with timely, relevant and accurate official statistics with related agricultural indicators.

The Rwanda experience also shows that geographic information systems (GIS) and information and communication technology (ICT) tools offer significant improvements for efficient data collection and management compared to traditional systems using paper-based questionnaires and printed maps.

Introduction

The National Institute of Statistics of Rwanda (NISR) is responsible for coordinating the national statistics system and for providing timely, accurate, and useful statistics for all economic sectors, including agricultural statistics. This is accomplished by coordinating, at the national level, efforts to collect, archive, analyze, document and disseminate data within an integrated and sustainable framework.

The agriculture sector is one of the most important economic sectors of Rwanda, accounting for 32% of the gross domestic product (GDP) and 70% of the work force. Since 2012, the NISR has embarked on designing and implementing a system of agricultural statistics that can provide information for the agriculture sector, such as crop production estimates, livestock inventories and basic social and economic data.
The first year of the Seasonal Agriculture Survey (SAS), first conducted in 2013, has enabled NISR to design the survey programme starting from the 2014 agricultural year. Three years later, at the end of 2016, an improved version of the survey system was put in place with the aim of improving NISR’s ability to respond to stakeholder requests and provide more dynamic indicators.

This paper includes a historical evolution of agricultural statistics in Rwanda, and contains a description of the agricultural statistical system used for the period 2014–2016. A detailed description of the upgraded SAS and related improvements are also included.

Seasonal Agricultural Survey for the period 2014–2016

General characteristics of the survey design

The overall survey design of the “Seasonal Agricultural Probability Survey Programme” was a seasonal, national, multiple-purpose, agricultural probability sample survey designed to obtain timely and reliable basic data for the agriculture sector at the national level. The survey design established was based on multiple frame survey methods. This section describes the survey design parameters and special characteristics, and the activities carried out for the establishment of the Survey Programme.

The decision to base SAS on multiple frame sampling methods took into account local conditions and resources, and alternative survey methods. Each agricultural year, the SAS Programme consists of three main seasonal surveys during early growing periods:

- Survey season A, phase 1 (area estimations): November–December;
- Survey season A, phase 2 (yield/production estimations): February–March;
- Survey season B, phase 1 (area estimations): April–May;
- Survey season B, phase 2 (yield/production estimations): June–July;
- Survey season C, (marshland): August–September;

For each of the two first surveys (seasons A and B), crop production estimates for some of the main survey variables were obtained, combining phase 1 and phase 2. The two follow-up surveys (season A, phase 2 and season B, phase 2) were undertaken using plots belonging to a subsample of resident farmers of the
area sample surveys conducted as phase 1. In all cases, field data collection was carried out by enumerators who interviewed the farmers directly.

The determination of the variables to be studied and their required level of accuracy were clearly established at the outset of the survey planning because this would have a direct bearing on the overall sampling design and, in particular, the type of multiple frame estimators used, the questionnaire design, and data collection procedures.

The multiple frame sampling methods applied combined a probability sample of land areas called “segments”, selected from an area frame, with a complementary short list of “special farms”. The multiple frame estimates combined estimates from the area sample with estimates obtained from the list of special farms.

In terms of resources required, nine people (SAS personnel) were involved at headquarters in Kigali, including the SAS coordinator (the Director of Economic Statistics). In addition, five NISR personnel in the cartography area were heavily involved. Conducting SAS also involved the work of 5 regional supervisors, 30 supervisors and 120 enumerators for field work and a budget of USD 1.3 million. The initial cost of acquiring the orthophoto imagery (25 cm resolution) was approximately EUR 107/km² and entirely covered by the Rwanda Natural Resource Authority (RNRA).

**List sample component of the multiple frame survey**

The complementary list of special farms ensures the inclusion of those farms that make a significant contribution to the total estimate of some important survey variables.

The definition of special farm was provided by agricultural experts and veterinarians in each district, using the following criteria: a farmer growing crops on at least 10 ha of land or any farmer raising 70 or more cattle, 350 goats and sheep, 140 pigs, 1 500 chickens, or managing 50 bee hives.

The list of special farms is updated once a year. At the last update, 499 special farms were considered. From those, only 20 had an intersection with sample segments.

Because the area frame includes all of the land in the country, larger special farms are included in the area frame. These large special farms are also included on the list frame so it is possible that their production may be included twice in
the dual frame design. It is, therefore, important to identify the duplication and remove it from the system design.

Cartographic work of delineating the boundaries of special farms inside the selected sample segments was absolutely necessary but was not a demanding job.

Special farms are treated separately in the survey design as data from farms on the list is at the farm level while data from the farms selected in area frame segments is for the farm tract (land inside the segment).

**Area frame construction and replicated area sample selection**

**Area frame construction**

**Agricultural strata:** The area sample design consisted of a stratified probability sample of segments, with a replicated selection procedure. From the improved stratification, the total land of Rwanda was subdivided into 12 non-overlapping strata as shown in Table 1.

<table>
<thead>
<tr>
<th>Stratum number</th>
<th>Stratum definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1</td>
<td>More than 50% cultivated area, Seasons A and B only</td>
</tr>
<tr>
<td>1 2</td>
<td>More than 50% cultivated area, Seasons A, B and C</td>
</tr>
<tr>
<td>2 1</td>
<td>Marshlands for other crops</td>
</tr>
<tr>
<td>2 2</td>
<td>Marshlands potential for rice</td>
</tr>
<tr>
<td>3</td>
<td>Rangeland (between 5% and 50% cultivated area)</td>
</tr>
<tr>
<td>4</td>
<td>Non-cropland</td>
</tr>
<tr>
<td>5</td>
<td>Cities and towns</td>
</tr>
<tr>
<td>6</td>
<td>Lakes and other water bodies</td>
</tr>
<tr>
<td>7</td>
<td>National parks</td>
</tr>
<tr>
<td>8</td>
<td>Uncultivated marshlands</td>
</tr>
<tr>
<td>9</td>
<td>Forests</td>
</tr>
<tr>
<td>10</td>
<td>Tea plantations</td>
</tr>
</tbody>
</table>
Sample selection

The area sample can be considered as a stratified, cluster sample of tracts, with a tract consisting of the part of a holding (or non-agricultural areas) included inside the segment. A replicated systematic selection method was used to select the segments. Among the 12 strata, only five were sampled, covering 17 596.20 km² because the other strata did not contain information relevant to the survey programme. Eighty-four percent of intensive agriculture is found in the first and second strata. These are key strata for the area frame construction and sample selection.

The strata, primary sampling units (PSUs), zones, and sample segments have identifiable physical boundaries (e.g. roads, paths, rivers) that can be located both in the field and on the cartographic materials used for their identification. For 2014, seasons A, B, and C, the PSUs were delineated having a total size between 100 ha and 200 ha.

Sample size: Rwanda's sample design has segments of equal target size in each stratum. As a result of experience in data collection, it was decided to delineate segments of approximately 10 ha in the sampling universe (originally segments of 20 ha were constructed). This was done to reduce the size of the cluster in the survey design. However, for the rangelands, due to the lack of physical boundaries for small areas, selected segments consisted of 50 ha.

The number of sample segments was determined by a large number of factors. For instance, the resources available, precision of data required, enumerators’ workload, and the required frequency of data collection: it is required to undertake 5 field data collection operations for each agricultural year to cover area and yield of three seasons. Therefore, it has been concluded, in view of the experience gained, that the largest possible sample size should be less than 600 segments.

Based on work previously done during the 2013 pilot survey, the total sample size was determined to be n = 540 segments in all strata (see Annex 3).

The sample allocation of the 540 segments in strata was highest in the intensive agriculture (seasons A and B) stratum (63%), followed by marshland for rice (12%). Segments were equally distributed in both the rangelands stratum and intensive agriculture stratum (season C plus seasons A and B). The distribution of segments was smallest in marshlands for other crops stratum (7%). In all cases, a tolerance interval for segment delineation of 20% was adopted.
The five sampled strata were subdivided into a total of 10,560 PSUs, and 540 PSUs were selected. The selection process is indicated in the section below. In the 540 PSUs, 8,556 segments were delineated and only 540 segments were selected for data collection.

**Sample selection methods**

For the first phase of the SAS survey for 2014–2016, the samples in each stratum were selected for representativeness at the national level. The area frame sample replicates were selected independently in each stratum. The same segments were visited in 2014, 2015 and 2016.

**Replicated systematic selection method**

Each stratum was subdivided into PSUs, which were ordered and listed in a serpentine fashion first by district and then within the district. Next, GIS software was used to measure the size of the PSU in hectares. This size was then divided by 10 ha for strata 1.1, 1.2, 2.1 and 2.2; and 50 ha for stratum 3.0, and rounded to the nearest integer. PSUs were then selected by probability proportional to size (PPS) – one segment was selected, following a systematic procedure in each selected PSU for the segment to be enumerated in the field.

**Field data collection**

Two field visits (two phases) were considered for seasons A and B, and one field visit for season C, for a total of five visits to the field during the agricultural year. As previously mentioned, an orthophoto, called segment photo (cf, Annex 5), was prepared for the selected segments for field identification and data collection. The segment photo included the segment boundaries. In addition, to assist data collection, a GPS (geographic positioning system) and PDA (personal digital assistant) were used.

**Seasons A and B**

During phase I of the survey, every selected segment was screened using a screening form. In this way, enumerators accounted for every plot inside the segment. All tracts (part of the farms included in the segment) were classified as either agricultural (cultivated land, pasture, and fallow land) or non-agricultural land (water, forests, roads, rocky and bare soils and buildings). In other words, during phase I, a complete enumeration of all farms having agricultural land, and operating within the selected segment, was undertaken by using a screening form. The screening form was used to collect data on area under crops, and crops
planted. All tracts and plots (fields) were indicated on the segment photo and sent to headquarters to provide a measurement of all plots and other areas in the segment, to be used during phase II field work.

The screening form included additional questions to obtain data on anti-erosion activities, and to know whether a particular crop corresponded to the same season when harvested. The screening form in Rwanda has been used much more heavily than in most countries, due mainly to the lack of reliable agricultural information and also to compensate for the lack of farmers’ capacity to provide accurate data.

Phase II of the survey was mainly devoted to the collection of data on demographic and social characteristics of the farmers, and to obtain information on crop production. Such data were collected through the application of a questionnaire to a sample (say 25%) of the farms partially or totally included in each selected segment.

Livestock data were collected only during phase 1 of season A.

For each of the two first surveys (seasons A and B), crop production estimates for some of the main survey variables were obtained, combining phases 1 and 2. The two follow-up surveys (phase 2, season A, and phase 2, season B) were undertaken using plots (inside and outside of selected sample segments) belonging to a subsample of resident farmers of the area sample surveys conducted as phase 1.

**Season C**

In Rwanda, there are two main agricultural seasons (season A and season B). However, there are parts of the country where season C has crops (with short vegetative cycles) and other parts of the country where production in season C is negligible.

Season C (dry season) plays a complementary role of lesser importance to help fighting against food insecurity. Therefore, it was considered useful to assess its contribution to agricultural products compared to seasons A and B, both for farmers’ income and for its contribution to food security.

For this reason, only three strata: 1.2, 2.1 and 2.2. (out of five agricultural strata) were covered for season C. This means that the 150 segments selected for the concerned strata were covered during the dry season survey. The main crops affected by the assessment under this season are those that do not last more than
three to four months and must be in place between late May and early June. Moreover, perennial crops are not affected because they are taken into account either in season A or B. The same applies to rice.

Data entry, processing and analysis, report and final publication of data

Data entry of the completed and checked questionnaires was undertaken at NISR offices by 20 trained staff members using CSPro software. For data processing, analysis and presentation, the SPSS software was used. After data collection and map digitization, data were entered in computers (using CsPro), and then edited, cleaned and summarized (using SPSS).

Standard statistical techniques were used for treating outliers and to make adjustments for missing data. The tables prepared were then inserted and explained in the report for each agricultural season.

The area frame estimator should not consider the tracts corresponding to farms of the list of special farms, in order to avoid duplications. Therefore, all farms with tracts included in the sample segments were compared with those on the list of special farms, in order to eliminate duplications from the area frame estimator. As previously mentioned, 499 special farms were considered for the 2014 phase 1, season A survey, and 540 area sample segments. Of those 540 segments, only 20 segments had overlap with special farms.

Because the list of special farms is completely enumerated, the variance of the multiple frame estimator is equal to the variance of the area frame estimator, but the coefficient of variation (CV) of the multiple frame estimator will always be less than or equal to the CV of the area sample estimator. Multiple frame surveys, therefore, obtain more precise estimates than area sample surveys for those variables partially accounted for in the list of special farms.

The survey results have been published on a regular basis since 2014. The methodology of the SAS is accessible to the public, and the database is accessible to external users on NISR’s website. However, data users made it clear that they have found it difficult to interpret survey results based on land use, and expressed their needs regarding the diversity of agricultural data at the household level.

In 2016, NISR started the process of upgrading the SAS to take into account the above-mentioned concerns and to improve the reliability of national-level estimates compared to previous SAS results. The upgrading process had the specific goal of achieving reliable district estimates to meet the needs of
stakeholders of more representative indicators at the local levels of administration. By implementing a complete reconstruction of the area sampling frame, NISR increased its ability to respond to stakeholder requests and to provide more dynamic indicators.

**Overview of the proposed upgrading of the Seasonal Agricultural Survey**

The upgraded agriculture survey is similar to the SAS for 2014–2016 design based on an area segment sampling frame and a list of special farmers. However, the 2017 upgraded SAS used a complete redesign of the area frame and sampling strategy that provides district-level estimates of crop area and production, as well as a set of household and livestock modules containing questions on key indicators and/or monitoring variables.

The objective for upgrading the survey is to provide not only accurate district and national estimates\(^{28}\) on key agriculture and land variables, but also household level information on indicator and/or monitoring variables of interest that are agricultural as well as socioeconomic, environmental, and food security oriented from one survey source. The most common survey variables include:

- Population involved in agriculture (e.g. basic demographic characteristics of agricultural households, farmer's household members working in the farm, hired workers on the farm, days of work);
- Land (e.g. land tenure, agricultural land, crop land);
- Agricultural inputs (e.g. labor, type and quantity of seeds, fertilizers and pesticides, machinery, equipment and agricultural buildings, use of seeds, use of fertilizers, use of pesticides);
- Agricultural practices (e.g. farming methods, land use, irrigation and source of irrigation water, drainage, anti-erosion activities);
- Crop production (actual crop yield of each crop, number of trees, use of production);

---

\(^{28}\) Here, it is important to recall that even though the NISR has gone from a national sample selection to a district sample selection, the national system is more efficient for national-level data. The district-level system is better if there is a need to modify a single district at a time.
Livestock and poultry inventories (e.g. number, type, age, sex, breed and use);

Livestock products (production of milk, meat, eggs, honey);

Costs of agricultural production and value of sales;

Sustainable agriculture; and

Food security.

In terms of coverage, the upgraded SAS design includes an increase in sample sizes, which provides the same national coverage of crops and livestock as the previous SAS, plus additional estimates of selected agricultural commodities, and indicator and/or monitoring variables with sufficient accuracy and precision for publication at the district level. A newly constructed area sampling frame provides coverage of all agricultural land plus households located in rural village communities. This area frame excludes urban areas, national forests, and non-agricultural land use such as local government buildings, shopping centres, warehouses and gasoline stations, even in rural communities.

The general approach to upgrading the SAS area frame segmentation is based on more efficient integration of agricultural land and associated farm households into a more consolidated and reduced number of land-use stratum for sampling at the district level.

The survey sampling frame design also includes the use of a large-scale farm (LSF) list frame that was previously used to supplement the area frame to prevent such large farms (if they fell within a sampled segment) to over-expand the estimates of their corresponding crop and livestock numbers.

The LSF is totally enumerated (a 100% sample), and survey data handling procedures require any LSF data found in a sample segment to be considered as duplicated data, and so removed (from the area frame) before summarization of the data for publication.

The new area sampling frame sampling units or segments were constructed by NISR staff and GIS experts with previous experience in processing and editing digital mapping imagery made available by RNRA.

The detailed methodology and implementation strategy are described in a complementary paper by NISR.
Annex 1

Land stratification

Map 1.1: Land stratification in SAS 2014–2016

Map 1.2: Improved land stratification for upgraded SAS 2017
Annex 2

Example of a primary sampling unit and breakdown into segments (2-m image)
Annex 3

Sample segments by strata

<table>
<thead>
<tr>
<th>Strata</th>
<th>SAS 2014–2016</th>
<th>Upgraded SAS 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stratum number</td>
<td>Number of Segments</td>
</tr>
<tr>
<td>Intensive agriculture (hillside)</td>
<td>1.1</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>48</td>
</tr>
<tr>
<td>Marshlands</td>
<td>2.1</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>40</td>
</tr>
<tr>
<td>Rangelands</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>Total area and number of segments</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>Stratum village</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Annex 4

Location of sample segments

Annex 5

Segment map for data collection

Map 5.1: Segment map as delineated in SAS 2014–2016

Legend
- Landmark
- Plot boundary
- Segment boundary

<table>
<thead>
<tr>
<th>Landuse</th>
<th>Area, Sq Meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>96,037.03</td>
</tr>
<tr>
<td>Non-Agriculture</td>
<td>5,923.35</td>
</tr>
<tr>
<td>Pasture</td>
<td>0.00</td>
</tr>
<tr>
<td>Grand Total</td>
<td>101,960.36</td>
</tr>
</tbody>
</table>
## Annex 6

### Seasonal Agricultural Survey Sample Details

<table>
<thead>
<tr>
<th>Seasonal Agricultural Survey (SAS) Sampling Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strata</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td><strong>Season A</strong></td>
</tr>
<tr>
<td>Phase 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Phase 2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td><strong>Season B</strong></td>
</tr>
<tr>
<td>Phase 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Phase 2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>
UNITED STATES OF AMERICA

By Sarah Hoffman

Introduction

Prior to 1997, the United States Census of Agriculture was conducted by the Census Bureau, within the Department of Commerce, using its own sampling frame. Agricultural surveys, including production surveys, were conducted by the Department of Agriculture (USDA), using its own frames. Responsibility for the Census of Agriculture was transferred to the Department of Agriculture in 1997, allowing USDA/National Agricultural Statistics Service (NASS) to use the same sampling frames for the census and agricultural surveys. USDA/NASS uses both a list frame and an area frame.

There are 3.2 million farmers operating 2.1 million farms in the US, which is only about 1% of the total population. To be counted as a farm in the US, an operation must produce and sell, or normally produce and sell, USD 1 000 in agricultural products. Every country should use a sampling frame(s) that meets their statistical requirements and their practical reality. USDA/NASS does not use a sampling frame that is directly tied to the population census, but rather uses a multiple frame approach for both the Census of Agriculture and many agricultural surveys.

USDA/NASS is not the only agency within the USDA to maintain a sampling frame. The Natural Resources Conservation Service uses an area frame for the National Resources Inventory survey. However, this paper focuses on the use of sampling frames by USDA/NASS for the Census of Agriculture and agricultural surveys for acreage, production, inventory, stocks, inputs, prices, expenditures and more.

Area frame methodology

The area frame used by USDA/NASS contains all of the land within the US, with the exception of Alaska. The area frame is stratified by land use, and uses segments of land as its sampling unit. The area frame is constructed one state at a time, one county at a time. The first step in area frame construction is to delineate land areas into land-use categories, using satellite imagery and other information, such as aerial photography and topographic maps. USDA/NASS has a digital system to create an area frame for each state, and uses satellite data.
for stratification and digital line graph data for boundary identification. Table 1 shows the area frame stratification for a typical state. Within a county, homogeneous blocks of land called primary sampling units (PSUs) are identified, measured, and assigned terminal sampling units until all land in each stratum in a county is completed. By measuring the size of all PSUs, the land area in each stratum is determined. The desired size of the PSU varies by strata, but the more intensively cultivated strata generally contain an average of 6–8 segments.

In most strata, the population is divided into equally sized subpopulations called substrata. Substratification ensures that the sample will be distributed throughout the state. It also improves the precision of the estimates because the substrata are designed to be more homogeneous in terms of the type of agriculture present.

Table 1. Area frame sampling in the United States, by stratum, for a typical state

<table>
<thead>
<tr>
<th>Stratum description</th>
<th>Segment size (square mile)</th>
<th>Number of segments</th>
<th>Number of substrata</th>
<th>Segments selected per substrata</th>
<th>Total segments selected</th>
<th>Expansion factor (sampling rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% or more cultivated</td>
<td>1.00</td>
<td>29 340</td>
<td>15</td>
<td>15</td>
<td>225</td>
<td>130.4</td>
</tr>
<tr>
<td>50–74% cultivated</td>
<td>1.00</td>
<td>10 020</td>
<td>6</td>
<td>10</td>
<td>60</td>
<td>167.0</td>
</tr>
<tr>
<td>15–49% cultivated</td>
<td>1.00</td>
<td>9 960</td>
<td>5</td>
<td>10</td>
<td>50</td>
<td>199.2</td>
</tr>
<tr>
<td>Ag-Urban</td>
<td>0.25</td>
<td>10 182</td>
<td>3</td>
<td>10</td>
<td>30</td>
<td>339.4</td>
</tr>
<tr>
<td>Residential/Commercial</td>
<td>0.10</td>
<td>12 792</td>
<td>2</td>
<td>10</td>
<td>20</td>
<td>639.6</td>
</tr>
<tr>
<td>Range</td>
<td>2.00</td>
<td>5 135</td>
<td>5</td>
<td>5</td>
<td>25</td>
<td>205.4</td>
</tr>
<tr>
<td>Non-agricultural</td>
<td>Varies</td>
<td>28</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>Varies</td>
</tr>
<tr>
<td>Water</td>
<td>-</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td>Water not sampled</td>
</tr>
</tbody>
</table>

Based on an analysis of previous survey data, an optimal sample allocation is generated at the national level to determine the smallest number of segments required to achieve a certain precision. Data collection consists of obtaining information for all land within the boundaries of the sampled segment. To associate agricultural commodities with a segment, it is necessary to establish reporting units:
Tracts

Each segment is divided into tracts of land, each representing a unique land operating arrangement. Crops and land located within the boundaries of a tract are associated with the tract.

Weighted Segment

Another type of reporting unit is the farm. A farm can be in one or more area segments that may or may not be in the sample. Livestock located throughout the entire farm can be linked to a segment using the weighted segment approach, which assigns livestock on an operator’s entire farm to a segment based on the proportion of the farm’s acres within that segment. An important requirement is that the sum of weights across all segments containing parts of the operator’s entire farm is equal to one.

Challenges

The benefit of an area frame is that it satisfies the requirements of a probability survey. It is complete and every acre of land has a known chance of selection. By adhering to the rules used to associate livestock with land, every animal has a known probability of being selected. The strengths of an area frame is that it includes complete coverage, longevity, and estimates well for very common commodities. Its weaknesses include cost (expensive to create, and data collection can be costly), high respondent burden, difficulty in targeting specific or rare commodities, sensitivity to outliers, and requires definable physical boundaries.

List frame methodology

USDA/NASS uses and maintains a list sampling frame of farm operators and agribusinesses. When a name is selected for a survey, the information requested pertains to crops and livestock on the total acres operated by the selected name. The sampling unit is a name but the reporting unit is all land operated by that name.

To avoid duplication and maintain the probabilities of selection, it is necessary to determine if the same units of land can be reported by more than one name on the list. All names associated with each unit of land must be obtained to determine if they are also in the sampling frame and have a chance of being selected. Survey rules and procedures must be followed to ensure that the probabilities assigned to each animal and crop acre are maintained and not duplicated.
A properly constructed list frame is an efficient sampling tool with several noteworthy strengths. Because names and addresses are known, cheaper methods of data collection such as mail (including sending instructions for self-reporting online) and telephone interviews can be used. Another strength of the list frame is the ability to target specific or rare commodities. Sampling efficiencies are gained when list commodity information is classified and names are stratified by size groups. Stratification also ensures that extremely large operations are properly accounted for. Table 2 shows an example of list stratification and sampling for a cattle survey.

Table 2. Example of list frame sampling, by stratum, for a cattle survey

<table>
<thead>
<tr>
<th>Stratum description</th>
<th>Population size</th>
<th>Sample size</th>
<th>Expansion factor (sampling rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–99 head of cattle</td>
<td>20 200</td>
<td>161</td>
<td>125</td>
</tr>
<tr>
<td>100–299 head</td>
<td>15 100</td>
<td>138</td>
<td>109</td>
</tr>
<tr>
<td>300–499 head</td>
<td>10 000</td>
<td>125</td>
<td>80</td>
</tr>
<tr>
<td>500–749 head</td>
<td>5 000</td>
<td>125</td>
<td>40</td>
</tr>
<tr>
<td>750–999 head</td>
<td>2 000</td>
<td>200</td>
<td>10</td>
</tr>
<tr>
<td>1,000 + head</td>
<td>300</td>
<td>300</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>52 600</strong></td>
<td><strong>1 049</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Challenges**

To satisfy the requirements of probability sampling, every animal and crop acre must be uniquely associated with one name on the list and each name must have a known probability of selection. The list frame does not satisfy the requirement of completeness (names are potentially missing). Farms go out of business, farmers retire, and new farms are started every day. Another related weakness is that it becomes out-of-date quickly. It also requires ongoing maintenance to build, update, remove duplication, and remove out-of-scope records.
Multiple frames

Methodology

Multiple frame sampling takes advantage of the best features of both the area and list sampling frames – the completeness of the area frame and the efficiency of the list frame. Other strengths include the ability to control costs (large list, small area samples), to target specific or rare commodities, and to control variability due to sampling. It relies on the joint use of the sampling frames and has only two requirements:

1. every element of the survey population must be included in at least one of the frames, and
2. the overlap of sampling units between frames must be determined to avoid duplication.

Because the area frame is complete, every animal and every acre of land has a chance of being selected in an area sample. Therefore, in a multiple frame survey, every area reporting unit must be checked to determine if it could also have been selected from the list frame. The non-overlap domain consists of land associated with an area sample that is not accounted for by the list frame because the area operator is not on the list and, therefore, did not have a chance to be selected from the list frame. The overlap domain consists of land that can be reported by both an area unit and a list unit.

The most critical aspects of multiple frame sampling are the unique association of names with land and the determination of overlap between frames. These aspects complicate survey procedures. Detailed information about the operator of the land, other names under which the operator does business, and other people associated with the operation must be collected. This allows the overlap between the area and list frames to be identified and ensures that every name is represented only once.

Depending on the needs of the survey, sampling procedures can differ by specific surveys. In general, USDA/NASS uses stratified sampling for the area frame, and either stratified sampling, probability proportion to size (PPS) or multivariate probability proportion to size (MPPS) sampling for the list frame. In all cases, replicated sampling can be used.
Estimators

Multi-frame estimators are the summation of list estimators with the area non-overlap estimators. The actual list estimators and area estimators used, and their corresponding variance estimators, depend on the item being estimated, the sampling techniques used, and the method of non-response adjustment used.

Multiple indications of the estimate can be generated using this multiple frame, replicated sampling approach. For example, for acres planted in specific crops, estimators can be calculated for the list frame only, area frame only, non-overlap domain, and ratio estimators comparing year-to-year or quarter-to-quarter. Because the official estimate represents a composite of information from various estimators (including remote sensing), direct sampling errors are not published. The relative standard error for the 2014 US corn acreage was 1.1%, soybeans 1.1%, winter wheat 1.9%, other spring wheat 3.8%, upland cotton 3.1% and sorghum 5.3%.

For the Census of Agriculture, the entire list frame becomes the census list. The area frame sample from the June agricultural survey is used to quantify the number and types of farms that are not on the census list. Any operation identified as non-overlap between the June agricultural list and the census list is also enumerated. Together the list and area frames allow for calculation of the coverage and misclassification errors in the census. The CV for the 2012 census for the number of farms was 1.6%, and for land in farms, 0.9%.

Challenges

The weaknesses of multiple frame sampling is that: 1) the non-overlap domain can be small, 2) overlap determination can be difficult, 3) errors in overlap determination can bias estimates, and 4) the two frames must be maintained independently.

Conclusion

USDA/NASS uses this multiple frame approach to estimate acreage, production, stocks, economic data, and more for many commodities. The multiple frame approach has served USDA/NASS well, and it continues to be at the heart of survey and census processes. It is a powerful and adaptable tool, although it requires a commitment to maintenance and overlap checking.
United States Department of Agriculture experiences in Nicaragua, Nigeria and Tanzania

by Sarah Hoffman

The underlying principle of a master sampling frame is that it is the source for all agricultural samples and censuses, and provides samples for demographic and household surveys. In many developing countries, no comprehensive source exists. The need for agricultural statistical data cannot wait until such a master sampling frame exists. Given a lack of a strong list frame, one alternative is to employ area frame surveys. The United States Department of Agriculture’s National Agricultural Statistics Service (USDA/NASS) has assisted, or is in the process of assisting, three countries — Nicaragua, Nigeria and Tanzania — in creating an area frame with point sampling in order to help meet their needs for agricultural production data.

This approach can aid a country in ultimately developing a master sampling frame because it begins with classifying the land cover of a country via satellite imagery. This paper looks at the development of an area frame, sample design, and other procedures used in these three countries, as well as further challenges.
**Nicaragua**

**Introduction**

USDA/NASS was requested to assist the Government of Nicaragua in developing an intensive and statistically sound crop and livestock estimating programme. Work was carried out in the 1990s with the collaboration of the Ministry of Agriculture, Central Bank of Nicaragua, United Nations Development Program, and the United States Agency for International Development. Previously, Nicaragua made agricultural estimates based on non-probability list samples. The quality and completeness of these list samples was unknown, and the quality of estimates derived from these surveys was suspect at best. To obtain a complete measure of total basic grain production at the national level and for 7 regions and 16 departments, it was determined that an area frame should be constructed to obtain reliable estimates.

**Methodology**

Area frame point sampling was chosen as the methodology. Satellite imagery would have been preferred but was far too costly at the time. Topography maps (scale 1:50,000) were used to stratify the land into six stratum (see Table 1). Standard deviation approximations were used to compute optimal sample allocation, with a sample size of 5,600 determined to be necessary to provide statistically valid estimates at the department level.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Stratum definition</th>
<th>Area (km²)</th>
<th>Manzanas (^1)</th>
<th>Sample allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>High cultivation 60–100%</td>
<td>12 400</td>
<td>1 178 492</td>
<td>2,844</td>
</tr>
<tr>
<td>II</td>
<td>Medium cultivation 40–59%</td>
<td>27 255</td>
<td>3 908 928</td>
<td>2,040</td>
</tr>
<tr>
<td>III</td>
<td>Low cultivation 1–39%</td>
<td>14 069</td>
<td>2 017 870</td>
<td>716</td>
</tr>
<tr>
<td>IV</td>
<td>Cities</td>
<td>259</td>
<td>37 240</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>Other towns</td>
<td>110</td>
<td>15 858</td>
<td>0</td>
</tr>
<tr>
<td>VI</td>
<td>Non-agricultural</td>
<td>16 439</td>
<td>2 235 805</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>70 532</strong></td>
<td><strong>9 394 195</strong></td>
<td><strong>5,600</strong></td>
</tr>
</tbody>
</table>

\(^1\) 1 manzana = 0.7 hectare = 1.73 acres
Random points were selected within the boundaries of each stratum. The enumerator then determined who operated the land where the point fell, and then enumerated only the one operator to determine agricultural activity on the entire farming operation. Occasionally, the selected random point fell in an area where the enumerator could not determine who operated the land (property boundaries, roads, streams). In these cases, enumerators were instructed to interview the farmer northeast of the selected point. If the sampled point fell in an urban area or other non-agricultural land, the sampled unit was treated as a zero report.

Enumerators were generally from the geographical area they would be enumerating and attended a three-week workshop. The first two weeks of training were devoted to properly locating the sampled point, while the third week focused on questionnaire content and interviewing techniques. Supervisors conducted quality control on approximately 20% of the assignments.

After field supervisors reviewed all questionnaires and sent them to the central office, the data were entered into a database and run through an edit. Data irregularities were reviewed and, if necessary, corrected until data were determined to be clean. Further analysis was conducted, such as reviewing listings of potential outliers for expanded data. The data were adjusted for nonresponse, although the response rate was 96%.

**Estimators**

The probability of selection is the size of the sampled farm divided by the land in the stratum of the department. That is, the probability of selection is proportional to the size of the farm (PPS). The sample was drawn with replacement (WR) PPS sampling. Data were expanded using the following estimator of population total, \( Y = y_1 + \ldots + y_n \):

\[
\hat{Y}_{WR} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}
\]

Where:

\[ p_i = \frac{x_i}{X} \]

\[ x_i = \text{total farm area for sampled point } i \]

\[ y_i = \text{survey value for sampled point } i \]
\[ X = x_1 + \ldots + x_n \]

This estimator is unbiased with variance:

\[ V(\hat{Y}_{WR}) = \frac{1}{n} \sum_{i}^{N} p_i \left( \frac{y_i}{p_i} - Y \right)^2 \]

An unbiased estimator of \( V(\hat{Y}_{WR}) \) is:

\[ \hat{V}(\hat{Y}_{WR}) = \left[ \frac{1}{n(n-1)} \right] \sum p_i \left( \frac{y_i}{p_i} - \hat{Y}_{WR} \right)^2 \]

**Results**

Below are some selected results from the 1995 Nicaraguan Agricultural Survey.

**Table 2. Selected results from 1995 Nicaraguan Agricultural Survey**

<table>
<thead>
<tr>
<th>Item</th>
<th>Direct expansion manzanas</th>
<th>Coefficient of variation (Cv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land in farm</td>
<td>7 913 484</td>
<td>0.7 %</td>
</tr>
<tr>
<td>Permanent cropland</td>
<td>1 255 042</td>
<td>1.9 %</td>
</tr>
<tr>
<td>Annual cropland</td>
<td>389 515</td>
<td>3.7 %</td>
</tr>
<tr>
<td>Pasture</td>
<td>3 848 361</td>
<td>1.0 %</td>
</tr>
<tr>
<td>Rangeland</td>
<td>1 571 233</td>
<td>1.9 %</td>
</tr>
<tr>
<td>Forest</td>
<td>705 899</td>
<td>2.7 %</td>
</tr>
<tr>
<td>Farmsteads</td>
<td>94 612</td>
<td>4.0 %</td>
</tr>
<tr>
<td>Other areas</td>
<td>48 821</td>
<td>8.3 %</td>
</tr>
<tr>
<td>Corn planted</td>
<td>353 916</td>
<td>3.1 %</td>
</tr>
<tr>
<td>Beans planted</td>
<td>86 257</td>
<td>5.6 %</td>
</tr>
<tr>
<td>Dryland rice</td>
<td>37 310</td>
<td>11.7 %</td>
</tr>
<tr>
<td>Sorghum planted</td>
<td>36 857</td>
<td>11.2 %</td>
</tr>
</tbody>
</table>
Future considerations

Although this was considered to be a successful project, if it were to be done today, satellite imagery would be used. It was recommended that replicates be used to incorporate a sample rotation scheme, as well as to allow other surveys to use the same sampling frame. By rotating the sample, some of the previous year’s samples are rotated out but most remain in the current year’s sample. This scheme was recommended because it reduces respondent burden and provides reliable measures of change through ratio estimators.

A major weakness of this area sampling frame is for items that are not proportionately associated with cultivated land use, such as specialty or rare crops. For such commodities, it is recommended that a list frame or multiple frame approach be used, with the list covering producers of rare crops, large crop producers, or large livestock producers.
Nigeria

Introduction

The purpose of the pilot for the area frame survey using point sampling was to study the practicality of using an area frame to estimate crops and livestock in Nigeria in order to introduce a statistically defensible methodology and produce more timely and accurate statistics. Prior to the pilot, Nigeria used a household-based agricultural product survey. In the agricultural product survey, households are selected, and for farming households, all farms cultivated by the selected household are measured. This can lead to a longer-than-desired sampling timeframe and burden on the household for reporting data. Some survey reports are released so late that they are not used by project sponsors and management, and primary stakeholders.

A point sample area frame could provide a quicker and easier way to select a statistically sound sample of farm operators than the household-based sampling methods currently used. Points were randomly located on the ground. Operators of the land at these points were interviewed, and data about their farm land and livestock were obtained.

The pilot study was conducted in Kaduna State and focused on wet season crops planted from May through August. This state was selected because: 1) it has small, medium and large farms; 2) a variety of vegetation and crops are grown there; 3) Agricultural Development Program management was supportive of the project; and 4) its proximity to Abuja allowed easy and relatively inexpensive oversight.

Output from this survey included State-level estimates of number of farms, land in farms, land areas planted to various crops and production of these crops, and inventory numbers for various livestock species and for cultivated fish. All of these items were calculated by type of farm – corporate farms and non-corporate farms.
Methodology

Frame construction

Spot 5 satellite images of 5-meter resolution were used in the area frame construction, along with geographical information system (GIS) layers for roads, railroads, rivers, political boundaries, and other features. The software used was ArcGIS. Six strata were used, all based on land use (see Table 3).

Stratification work was done one local government area (LGA) or a contiguous group of LGAs, at a time. Strata blocks were drawn using physical boundaries as guidelines, stopping at LGA boundaries. These blocks were dissolved into strata prior to sampling. The land area within each stratum block was measured and entered into an area frame database. The area frame database contained the block identification number, stratum code, and the land area within the block.

Sample selection procedures and size

Sampled points were allocated to strata as shown in Table 3. Points were randomly located within each stratum to ensure geographic distribution in all areas of the state. Each selected point was plotted on state and/or LGA maps for use in making enumerator assignments. Once assignments were made, the enumerator information was added to the sample database, along with geographic codes. Maps of assigned points were then generated for each enumerator. Global positioning system (GPS) coordinates for the sample points were given to the enumerators.

A sample of 600 points was selected. Ideally, estimates should have coefficients of variation (CVs) of less than 10% for items of major importance. Because this was a new technique for Nigeria, it was difficult to predict which CVs would be obtained. If farm data are likely to be highly variable, the sample size to assure 10% CVs would need to be very large. Budget limitations on the pilot survey could produce higher-than-desired CVs. The sample was allocated to strata in such a way that the number of points that actually identified farms were maximized. Stratum definitions and allocations are shown in Table 3.
Table 3. Stratum definition and sample allocations for Nigeria

<table>
<thead>
<tr>
<th>Stratum definition</th>
<th>Total land (ha)</th>
<th>Percent of land</th>
<th>Sample allocation</th>
<th>Percent of sample</th>
<th>Expected expansion factor¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural land, &gt;50% cultivated</td>
<td>2 169 389.62</td>
<td>49.1%</td>
<td>350</td>
<td>58.3%</td>
<td>1,240</td>
</tr>
<tr>
<td>Agricultural land, 15–50% cultivated</td>
<td>1 001 937.45</td>
<td>22.7%</td>
<td>150</td>
<td>25.0%</td>
<td>1,336</td>
</tr>
<tr>
<td>Urban/developed areas</td>
<td>33 484.03</td>
<td>0.8%</td>
<td>20</td>
<td>3.3%</td>
<td>335</td>
</tr>
<tr>
<td>Agricultural land, &lt;15% cultivated</td>
<td>1 171 074.26</td>
<td>26.5%</td>
<td>80</td>
<td>13.3%</td>
<td>2,928</td>
</tr>
<tr>
<td>Non-agricultural</td>
<td>36 052.34</td>
<td>0.8%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water (minimum 1 km²)</td>
<td>10 126.90</td>
<td>0.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Based on average size of farming operation = 5 hectares

**Estimators**

Data were expanded according to the following formulas.

\[ y_{ij} = e_{ij}p_{ij}x_{ij} \text{ and } \hat{y} = \sum_i \sum_j y_{ij} \]

Where:

- \( i \) = land use stratum,
- \( j \) = selected point within stratum,
- \( e_{ij} \) = expansion factor for selected point \( j \) within stratum \( i \),
- \( p_{ij} \) = population indicator for point \( j \) in stratum \( i \), and
- \( x_{ij} \) = survey value for point \( j \) in stratum \( i \).
Expansion factors are point specific, and were calculated as follows.

\[ e_{ij} = \frac{L_i}{n_i l_{ij}} \]

Where: \( L_i \) = total land in stratum \( i \),

\( l_{ij} \) = total land in farm identified by point \( j \) in stratum \( i \),

and

\( n_i \) = total number of points sampled (or usable) in stratum \( i \).

The population indicator was used to calculate expansions for different types of operations – all farms, corporate farms, non-corporate farms, and others. The value of the population indicator was “1” if the operation satisfied the definition of the population for which the expansion was calculated, and “0” otherwise.

The survey value depended on the item to be estimated. For total farm land, crops and livestock, the survey value was the total for the farming enterprise as reported by the farmer or measured by the enumerator. For estimating the number of farms, the survey value was “1”.

Variances, standard errors and CVs were calculated using the following formulas.

\[ V(\bar{y}) = \sum_i \frac{n_i}{n_i - 1} \sum_j (y_{ij} - \bar{y}_i)^2 \quad \text{where} \quad \bar{y}_i = \sum_j \frac{y_{ij}}{n_i} \]

\[ SE(\bar{y}) = \sqrt{V(\bar{y})} \]

\[ CV(\bar{y}) = \frac{SE(\bar{y})}{\bar{y}} \]
Data collection

Data collection was done in multiple phases. In late May and early June 2010, enumerators located the sample points and the associated operator. In June and July, they collected crop area information for the area where the point fell as well as for the whole farm, and laid out triangles for an objective yield measurement (cutting and weighing crop production within the triangle). From August through November, as the 2010 wet season crops were harvested, they obtained crop production data for the whole farm, including the objective yield measurement. Livestock, poultry and fish data were collected with a reference date of 1 October 2010. All questionnaires had to be returned no later than 15 December.

Enumerators were trained in both the classroom and the field on data collection procedures. An enumerator instruction manual was created for their reference and instruction. The manual had detailed instructions on the purpose of the survey, respondent confidentiality, sampling, terms and definitions, using the GPS devices (Garmin GPS Map 60CS), special rules for special situations, detailed question-by-question instructions, measuring the plot areas using GPS, locating and laying the triangles in plots for objective measurements at harvest, supervisor responsibilities and quality control, and a random number table for use in selecting the random location of the triangle.

Some of the rules for special situations included a point falling on the boundary of two farms or a farm and a non-agricultural land use, multiple family dwellings, a non-agricultural business where the operator also has a farming operation, or a farming operation with partners.

For the objective measurements, the random number tables and the length and width of the plots (measured by paces) was used to determine the random placement of the triangle. The area within the triangle was 1/100 of a ha. Some crops are harvested multiple times during the growing season, and required multiple measurements of the crop within the triangle.

One or two assigned points for each enumerator were randomly selected for quality control by their supervisor or office staff. The location of the selected point was verified. The operator was re-interviewed with a subset of questions from the original questionnaire.

When questionnaires were returned to the office, they were checked and reviewed for completeness. If necessary, they were sent back to the field to be completed. The data were entered in a computer database, where the computer
edit programs were run, and data were validated or updated as necessary. The computer edit checked for land areas and livestock items summations and basic logical relationships.

The questionnaires contained basic land-use questions. If a point fell on a farm or farmer’s residence, the crops planted to the field (and corresponding m² area) with the selected point were reported. The type of farm (legally registered or not) was reported. If the point did not fall on a farm or farmer’s residence, the type of non-agricultural land use was reported. Whole farm information was also collected, such as tenure, total area, area and production by crop for 15 specific crops, area of other crops, livestock inventories, poultry inventories, and fish inventories.

A computer program was developed to calculate expansions, variances, standard errors and CVs for all land, crop and livestock items. These items were generated for each strata and each type of farm as well as for the state.
Outcome

In the beginning, this project was a close collaborative process between Nigeria and USDA/NASS. USDA/NASS staff helped Nigerian staff create the area frame for Kaduna State, design the questionnaire, select the sample, prepare data collection materials, and train the trainers. Enumerator training was conducted in mid-May with data collection activities beginning immediately thereafter. Funding for this project ran out in May 2010. As a result, USDA/NASS staff were unable to return to Nigeria to observe data collection and to continue providing assistance for data processing and analysis as planned. Nigerian partners sent periodic status reports, the last of which indicated that all sample points were found, but not without some problems. From a distance, USDA/NASS was unable to troubleshoot or to provide assistance for resolving problems and issues.

With new funding obtained for 2011, NASS consultants returned to Abuja in January and March 2011 to provide hands-on assistance with data analysis, and document lessons learned from the pilot and develop plans for the way forward. The results were better than expected, but the pilot uncovered minor stratification issues, as well as some awkwardness in the questionnaire and instructions, all of which could be corrected with minor modifications. Survey results indicated that a slightly larger sample size will be needed to produce reliable statistics for the top ten crops produced in Kaduna State.

Nigerian partners wanted to repeat the pilot in Kaduna State during 2011 but frozen Government of Nigeria budgets did not allow funding in time for a May launch. Plans for 2012 had anticipated repeating the pilot in Kaduna and expanding to an additional state to test the methodology in a different agricultural situation. These pilots would have allowed for fine-tuning of the sampling frame, sample allocation, questionnaire and instructions, as well as development of information technology processes and programs for capturing, editing, analyzing and summarizing data that were not developed in 2010 due to the break in assistance. At this point, due to financial, security and political constraints, all plans are on hold.
Tanzania

Introduction

Tanzania, like Nigeria and Nicaragua, does not currently have a master sampling frame available for agricultural production surveys. It would like to improve its agricultural land and production statistics but does not have a well-defined list frame for its population of farms. In the past, the agriculture module in the population census was given only to a subsample of the population. For the 2012 census of population, the entire population census was assigned to the agricultural module. However, only aggregated data (number of households, number of population involved in agriculture) were available for evaluation as a sampling frame for agricultural production surveys. The point sample design was chosen prior to the availability of the 2012 population census data being available.

Methodology

The design has not yet been finalized, but will be similar to the one used in Nigeria. LandSat imagery with a 30-m resolution will be used. It is desirable to avoid the labor intensive manual delineation of strata and replace it with software that can derive and incorporate land cover maps. Currently, the plan is to use supervised classification, where the software gets “trained” to assign strata based on known land classification. Research is ongoing to review other land-use classification procedures to determine if there is a more efficient and cost-effective classification process. The sample design will be a stratified probability proportional to size sample, with the land stratified based on cultivated land. The strata will most likely be defined similarly to the Nigeria stratification plan, however, the sample sizes are still being discussed.

The cognitive pre-test conducted in November 2013 obtained data from sample points selected from land-use strata. For the pilot test conducted in 2014, the sample points from the land strata will have an enumeration area (EA) GIS overlay from the 2012 population census to identify the EA where the point falls.

The estimators will most likely be similar to the Nigeria point area survey.
Future considerations

Because this project has not been fully implemented, there are no known outcomes yet. One cost is fully known – LandSat imagery is free. It is believed that, given the stability and support from the Tanzanian government and other developmental partners, this project will be fully implemented, including post-survey analysis.

Conclusion

In all three countries, the main goal was to create or improve their agricultural production statistics, and not to create a master sampling frame. In Nicaragua, at the time the project was implemented, there was no conversation about a master sampling frame. In Nigeria, the entire project became too unstable to move towards a master sampling frame. In Tanzania, however, there can, and should, be movement towards a master sampling frame. Identifying the EA for each sample point will be a first step in attempting to tie the agricultural survey data in with the census of population data.

A point area sampling frame is a step towards a master sampling frame, with GIS identifiable points, data and farms. The sampled points in Tanzania can be mapped to an EA. Future challenges will include reconciling any definitional differences between the population census and its households, and the point area sampling frames points and farms. During the last population census, everyone was asked an agricultural module, but we have only seen basic information such as number of households and people involved in agriculture at the EA level. While it would be wise to maintain this information at a more detailed level, it would be even better to maintain a measure of size of agricultural production at a detailed level for efficient sampling of agricultural surveys.

Further considerations

Area sampling frames

We recommend using replicated sampling for area frames. Replicated sampling involves selecting several small samples instead of one large sample. For example, a replicated sample of 100 from a population of 1000 might involve selecting 5 replicates of 20 units each. There are multiple benefits to using replicated sampling, including reduced respondent burden, capturing changes from cycle to cycle, and cross-survey linkages.
Replicated sampling allows for rotation of samples for surveys done on a consistent cycle. In the above example, a sample rotation of 20% would equate to one new replicate every survey cycle. For an annual survey cycle, each year’s sample would be 80% (or four replicates) from the previous year’s sample and 20% (or one replicate) would be new. This reduces respondent burden and yet maintains enough of the sample to measure cycle-to-cycle change. The cycle-to-cycle change can be captured with ratio estimators.

Replicated sampling can also tie-in different surveys together (assuming they have a similar sampling scheme) while reducing respondent burden. For example, if you want to connect Survey A and B, but Survey C can stand alone, the use of replicates will allow for that and reduce the reporting burden for those sampled. In this example, replicates 1 and 2 could be used for Survey A; replicates 1 and 3 could be used for Survey B; and replicates 4 and 5 could be used for Survey C.

**Multiple frames**

Area sampling frames have a major weakness for items that are not proportionately associated with cultivated land use, such as specialty or rare crops. For such commodities, it is recommended that a list frame or multiple frame approach be used. Similarly, if there are relatively large operations that have a large impact on the item of interest, an area frame alone may be not suitable. A multiple frame approach, with a list of large and specialty operations and an area frame, would be more appropriate. This is particularly true in a country with mostly small household types of agriculture but with some large agricultural operations.

**Master sampling frame – household and agricultural production**

For production agricultural surveys, using measures of size of production makes for very efficient sampling. To utilize the list data from a population census as the master frame, there needs to be an agricultural section where the presence of, or actual values for, land, cultivated land and livestock inventories are captured. Capturing only the presence of agriculture in a household can provide the basic frame for a census of agriculture. Once the census of agriculture, with its actual values for land, cultivated land, area of crop and livestock inventories, is conducted, it can be an effective frame for sampling agricultural production surveys (preferably with a way of maintaining and updating the list between agricultural census cycles).

Relying on the presence of agriculture household information only, such as measures of the number of households involved in agriculture, is not efficient
for agricultural surveys. Although many countries have mostly small household types of agriculture, they may not have identical agricultural production. Assuming household information is sufficient for agriculture production assumes that there is little variance of area, crops, and livestock production across households. This is often not the case, as the number of farm operations and their associated production fluctuates within and across households.

It is recommended that measures of agricultural production (e.g. areas of cropland, total land and specific crops; number of livestock head) be connected to the sampling unit for agricultural surveys. If the population census serves as the initial source of the master list, this requires that agricultural modules on the population census include more than presence or absence of agriculture. The modules would need to collect a measure of size of agricultural productivity.