DEVELOPMENT OF MULTIPLE SAMPLING FRAMES FOR AGRICULTURAL SURVEYS IN KENYA

CROSS-THEMATIC SET F: Data sources / Data collection / Use of IT tools / Data quality

THEME TITTLE: Master sample frame for agricultural surveys

ABSTRACT

Agriculture is the mainstay of the Kenyan economy contributing about a quarter of the Gross Domestic Product. The country covers an area of 591,971 square kilometers. Eighty per cent of the land area is arid or semi arid and largely suitable for pastoralist activities. The balance is used for crop production and livestock keeping by the sedentary population. Despite the central role played by the sector in the economy, the quantity and quality of agricultural statistics in Kenya has worsened over time due to a myriad of factors. To address these challenges, the Kenya National Bureau of Statistics (KNBS) in conjunction with stakeholders in the Agricultural Sector and with support from the World Bank is in the process of developing a multiple sampling frame, a precursor to a comprehensive survey programme to address the identified data gaps. The multiple sampling plans will entail the construction of an area frame for collecting smallholder agricultural data and a list frame for collecting data from commercial farms. This approach is motivated by the fact that the smallholder farmers are very many and with small parcels of land while commercial farms are few and hence can be completely enumerated.

The above sampling strategy envisages a survey plan where agricultural input and output data is collected on a seasonal basis to obtain annual agricultural statistics using the area frame. An annual census of commercial farms and green houses will complement data from the smallholder farmers so as to obtain the overall picture of the performance of the agricultural sector on an annual basis. This approach which is taking place in Kenya for the first time is using a collaborative approach which involves the national statistics office as the lead agency, the Ministry of Agriculture, livestock and fisheries and various Non Governmental organizations pooling their expertise and resources together to generate the required infrastructure. This approach is preferred due to the positive aspects of institutional ownership, cost saving and the high quality of the resultant agriculture statistics that will be produced. This paper outlines the processes and sampling plans that will be undertaken to ensure that the quality of Agriculture statistics is of high quality. The International Conference on Agricultural Statistics (ICAS VII) forum is very timely as it will provide an excellent forum for validating and/or critiquing the entire process that Kenya is taking to address the quality of its Agriculture Statistics whose quality has been declining over time.

Key words: sampling frame, seasonal surveys,

Patrick M. Mwaniki
Senior Manager, Agriculture and Livestock Statistics,
Kenya National Bureau of Statistics, Nairobi, Kenya
P.O Box 30266,

Lt. Tumbo Street
00100, Nairobi, Kenya
Email: pmwaniki@knbs.or.ke

John G. Mburu
Manager, Agriculture Statistics,
Kenya National Bureau of Statistics, Nairobi, Kenya
P.O Box 30266,
Lt. Tumbo Street
00100, Nairobi, Kenya
Email: jmburu@knbs.or.ke

INTRODUCTION

Agriculture is the mainstay of the Kenyan economy contributing about a quarter of the Gross Domestic Product (GoK, 2016). The country covers an area of 591,971 square kilometers. Eighty per cent of the land area is arid or semi-arid and largely suitable for pastoralist activities. The balance is used for crop production and livestock keeping by the sedentary population.

Despite the central role played by the sector in the economy, the quantity and quality of agricultural statistics in Kenya has worsened over time due to a myriad of factors. To address these challenges, the Kenya National Bureau of Statistics (KNBS) in conjunction with stakeholders in the Agricultural Sector and with support from the World Bank is in the process of developing a multiple sampling frame, a precursor to a comprehensive survey programme to address the identified data gaps.

The multiple sampling plans will entail the construction of an area frame for collecting smallholder agricultural data, a list frame for collecting data from commercial farms and deployment of aerial surveillance methodology for pastoral areas. This approach is motivated by the fact that the smallholder farmers are very many and with small parcels of land while commercial farms are few and hence can be completely enumerated. A census of commercial farms and greenhouses is currently being undertaken countrywide.

METHODOLOGY

The Kenya Seasonal Agricultural Survey (SAS) is intended to inform policy decision and international reporting requirements. The approach adopted in this survey is based on a similar approach learnt from Rwanda (NISR, 2014). Compared to Rwanda, Kenya occupies a much larger geographical area and has a much higher smallholder farmer population. This therefore makes it even more critical to develop as area frame to facilitate periodic smallholder agricultural data collection. It is expected that data collected will be sufficient to generate county level estimates so as to support the data needs at that level. The survey will be conducted twice a year to cover both the long and short rain seasons. The pilot survey is expected to take off in April 2017 and then the main survey in October 2017 to coincide with the short rains season. This section outlines the methodology that will be used to develop the Area Frame and conduct the first round of the Seasonal Agricultural Survey (GoK 2015).

STRATIFICATION

Base Map Layer

The 2010 Land Use Land Cover (LULC) layer from KFS (Kenya Forest Service) will be used as a base layer and updated to come up with a current stratified layer of 2015. This layer was derived from ALOS (Advanced Land Observation Satellite) images which have a 10 Meters resolution. The said images did not have much cloud cover and where there were clouds DMC (Digital Monitoring

Constellation) 22Meters and LANDSATS 30 Meters images were used as a substitute. Higher resolution images will be used in the updating of the 2010.

Stratification
Definition of Strata
1. Built Up Area (Any artificial hard surface that is manmade)
2. Forest (Areas that have trees or shrub land)
3. Uncultivated Wetland (Areas that have water more than six months in a year but do not have any croplands)
4. Waterbody (Areas having water)
5. Bare Areas (Rock outcrops, Bare soil with no definite pattern)
6. Cropland (Areas where crops are planted and cultivated)
   • Small scale/Intensive agriculture below 20 ha
   • Large scale agriculture 20 ha and above
7. Pasture land (Areas used for animal forage)
   • Leys (Having clear fenced areas, regular shaped, established pastures)
   • Natural Grassland (Continuous fenced grassland areas)
8. Grassland (Areas covered with grass that are neither rangeland nor pasture)
9. Rangeland (Free range grassland and/or shrub land in Parks, reserves, conservancies, ranches)

The 2010 base layer covers the whole country with the classes being; Forestland, Cropland, Settlement, Wetland, Grassland and Other land. Assessment of the same led to the decision on subdividing the LULC to different Strata in a three level step: Primary, Secondary and Tertiary which will cover the whole country. Secondary strata will be a subdivision of the Primary strata while Tertiary strata are a further subdivision of the Secondary strata.

Imagery
Images (High Resolution Images, Aerial photographs and Google Earth) License - High Resolution Satellite Images. The high resolution satellite images will be procured by purchasing a license which would give access to parties involved to the full archive of images from 2002 to date. This purchase of the license will be cheaper than purchase of the images. The DRSRS (Directorate of Resource Surveys and Remote Sensing) is expected to provide Aerial Images of the same resolution as the satellite images in order to compliment the satellite images in the high potential agricultural areas. Google Earth images will not be used since we are having high resolution images already. If used the challenge would be reduced efficiency. The images in Google Earth are a mosaic of images from different years hence a reduction in accuracy. Shifts in the google Earth image products would have errors trickle down to the final Strata if used. Lastly, their use in fieldwork leads to an error of 10 meters due to registration. Based on the comparisons above, commercial software will be used in the stratification process, PSU as well as SSU creation. It should be noted that the organizations represented in this the Technical Working Group already have licenses hence no additional costs will be incurred in the use of the commercial software. However, open source such as QGIS may be utilized in field map production.

Ground Truthing Strata
Ground truthing exercise is guided by the need for clarification. Since the initial interpretation and automatic classification will be based on Satellites imageries, this requires validation using ground sample points. During the stratification process areas that are difficult to interpret or having conflicting categorization by the TWG are noted down. These areas will form a basis for verification. The sample points generated will not only clarify the areas but also act as points of assessing the accuracy of the Stratification process. The ground sample points will be uploaded to the tablets for use in the field. The sample points will be required to be in the same reference datum as satellite imageries for the purpose of compatibility. As indicated in the Table 1 above. During this process clarification will be on the tertiary level of stratification since that is the basis of the entire sampling frame. The task above will be undertaken by RS/GIS specialists from the organizations involved (RCMRD, KFS, DRSRS, NEMA, KALRO among others). Drivers will be sourced from the institutions mentioned above. Security personnel will be hired on a need basis. For the ground truthing exercise to be undertaken, Mobile Mapper Licenses will be purchased and

installed in the Tablets. The use of the mobile mapper stems from the fact that GPS would capture data as is while the mobile mapper will enable taking of coordinates and the ability to transfer the data to the server with the aid of good internet connectivity. The use of tablets will do away with cameras, GPS and batteries reducing the cost. The same tablets used in ground truthing the strata will be used during data collection. The full list of items required during ground truthing is indicated in the budget. Field sample points used in ground truthing will be based on first draft strata. Sample points will be randomly generated to verify the accuracy of the stratification. The duration of the ground truthing exercise for the whole country will take approximately 12days comprising of 10 teams.

PRIMARY SAMPLING UNITS
Shapefiles (Administrative Units, Rivers, Roads, Rail)
The primary sampling units will be required in a digital GIS format i.e. a shapefile. These are vector data sets that will be overlaid with the images for the purpose of analysis. The Kenya National Bureau will provide boundary data at National, County, Sub county, Locational and sub- locational levels. Kenya Agricultural and Livestock Research Organization (KALRO) will provide the Agro-Ecological Zones (AEZ). The above shapefiles will assist in locating both the administrative units and the Agro-Ecological Zones of various crop in the sampled holdings

Creation of PSUs
Primary Sampling Unit refers to the sampling units in the first stage of selection in a multistage sampling process. The output stratified layer in the previous process will form the base layer. Once the natural boundaries have been identified and extracted, the later will be used in the subdivision of the strata to form the PSUs. The PSUs will be developed by strata and a sample of the PSUs obtained through Probability Proportional to Size (PPS) sampling method. These PSUs will be subjected to an area rule whereby the PSU segments will be divided into blocks of 200 to 400 hectares. The sample size will be such that the estimates will be representative at the county level.

SECONDARY SAMPLING UNITS (SSUs)
Creation of SSUs
Secondary Sampling Unit refers to the sampling units in the second stage of selection in a multistage sampling process. The PSUs generated in the previous process will be used in the creation of the SSUs. These PSUs will be subjected to an area rule whereby the PSU segments will be divided into blocks of 10 hectares which are herein referred to as SSUs. Sampling of the SSUs will be based on systematic sampling method. The SSUs will have unique identifiers in the attribute table. The ID will be used in the selection of the samples.

ENUMERATION SURVEY
Map Production
To improve on the stratification accuracy, field maps will be produced for use through RS/GIS approaches. The maps will be used in validation of the stratification products. For every sampled segment, a segment image map will be produced to aid the enumerators in collecting data. A suitable size and scale of the map will be determined and cartographic visualization applied. The number of maps produced will be dependent on the coverage of the segment and crop calendar (seasons).

Ground-truthing

Field verification will be carried out for the selected SSU. The selected SSUs will have IDs identifying their boundaries. During field verification, the selected SSUs will be identified using landmark features such as boreholes, hospitals, schools, rivers or roads as indicated in the field maps.
 The following steps will be adopted for undertaking ground-truthing:-
Preparation of field questionnaires (check-list)
The TWG will determine which thematic areas before developing the questionnaires. These questionnaires will be in digital format to enable real-time transmission to the server. Hard copy questionnaires will be used as back up.

All land parcels in the SSU will be measured using tablets. This will be done by tracking of all the plots in the segment.

Data Collection

The team will be expected to collect various types of data to enable them enrich the initial available auxiliary data. Standard procedures will be adopted as follows:-

- Mobilization of a team of experts - The TWG will ensure that a team of experts are in place before the exercise takes off
- Formulation of a check-list - The check-list will ensure that data gaps are eliminated. This will minimize errors
- Mobilization of survey tools and equipment - Acquisition of tablets, vehicles, maps and items mentioned in the budget
- Publicity - The survey exercise has to be publicized before commissioning. This will enable smooth operation at the community level
- Fieldwork - The experts will be dispatched to different regions within the country based on the classes which will require validation.

Data cleaning, Analysis and Storage

Data from the field will be downloaded from the servers at the KNBS. Data from the server will be cleaned and weighted. Thereafter statistical tables will be generated for both seasons. SPSS software will be used in processing the statistical tables. After the analysis and production of the indicators, the output will be linked to their exact location using the unique ID on a GIS platform. The unique ID will be used to link the non-spatial and spatial data to complete the attributes in the spatial files. A Geo-data base will be designed to facilitate the linking and storage of the files. In the geo-database, access protocols are to be defined to allow access, editing and dissemination.

Awareness/publicity

To ensure maximum cooperation and participation from the citizens, publicity of the survey will be done using fliers, print, electronic media, barazas and other suitable channels. To enhance awareness creation and survey publicity, a team will be created comprising of KNBS and MOALF knowledgeable of the subject matter to steer the campaign through sensitization to the chiefs, village elders, sub chiefs and opinion leaders.

Training (TOT, Supervisors and RAs)

Two levels of training will be undertaken for the Pilot and the Main survey. These are:
Training of Trainers (ToT)
Training of supervisors and research assistants
The TWG will mount training of the TOTs who will then train the supervisors and the research assistants. The training will focus on:

- Mapping, ground truthing
- Use of tablets, mobile mapper
- SAS Questionnaire administration

Report Writing and Publication

The following reports will be prepared.

- Pilot / Pretest Report
- The Seasonal Agricultural Survey (SAS) and other Publications

The experiences and lessons learnt from the pilot report will be used to fine tune both the various survey instruments in the main survey as well as the mode of operation. The survey results and indicators will be published and disseminated to the stakeholders.

Dissemination

Dissemination of the results of SAS is important because it provides information for decision making to users such as policy makers, researchers and the general population among others. Some modes of dissemination include:

- Hold a workshop to release the results
- Highlight in the electronic and print media
- Upload in websites
- Distribute hard copies

REFERENCES

KNBS, (2016) "Economic Survey", Kenya National Bureau of Statistics, Nairobi, Kenya.
KNBS, (2015) "National Seasonal Agricultural Survey", Kenya National Bureau of Statistics, Nairobi, Kenya (Unpublished).
NISR., (2014) "Seasonal Agricultural Surveys", National Institute of Statistics, Kigali, Rwanda.