# Pilot Point Sample Area Frame for Agricultural Statistics in Nigeria

**Sarah Hoffman, Michael Steiner**
**U.S. Department of Agriculture-National Agricultural Statistics Service, Washington DC, USA**
**Corresponding author: Sarah Hoffman, e-mail: sarah.hoffman@nass.usda.gov**

A pilot agricultural survey was conducted in Kaduna State, Nigeria in 2010 to study the practicality of using area frame point sampling methodology to estimate crop acreage and livestock inventory in order to introduce statistically defensible methodology and produce more timely and accurate agricultural statistics. This pilot point sample area frame survey was designed and conducted in 2010 as a collaborative effort by the United States Department of Agriculture's National Agricultural Statistics Service (USDA-NASS), the Nigerian National Bureau of Statistics (NBS), the Nigerian Federal Ministry of Agriculture and Water Resources (FMAWR) and the Kaduna State Agricultural Development Project (ADP). As is common in many parts of the world, logistical and technology constraints as well as limited resources hindered efforts to improve the timeliness and accuracy of Nigeria's agricultural statistics. The goal of this project was to more efficiently use the resources available to improve the agricultural statistics in Nigeria. This process included developing an area frame, survey questionnaire, enumerator instructions and training, data collection and analysis. This paper looks at the sample design and other procedures used, problems encountered, and the outcome of this pilot point sample area frame survey.

Keywords: area frame design, agricultural survey questionnaire

## I. Introduction

There is a lack of timely and accurate agricultural statistics for Nigeria. This is caused in part by a lack of adequate resources, technology and organizational coordination of data collection efforts. By the Statistics Act of 2007, the National Bureau of Statistics (NBS) has the mandate for all official statistics in Nigeria and is responsible for providing accurate and timely information for policy makers. The Federal Ministry of Agriculture and Water Resources (FMAWR) recognizes that poor agricultural statistics impede planning and assessment of food security in Nigeria. NBS and FMAWR propose to develop appropriate interventions to improve Nigeria's agricultural statistics system.

As part of the U.S. Government's Global Food Security response for Nigeria, two consultants from USDA/NASS visited Nigeria in May 2009 to conduct an assessment, and identified activities for further collaboration to enhance Nigeria's agricultural statistics program. Nine Nigerian agricultural statisticians participated in a study tour coordinated by USDA/NASS in Washington, D.C. July 12-26, 2009. Team members were drawn from NBS, FMAWR, the Central Bank of Nigeria (CBN), and four State Agricultural Development Projects (ADPs). The two week workshop afforded participants the opportunity to undertake a case study of the U.S. agricultural statistics system, and learn about organization, methodology and management of programs for preparing official agricultural statistics. At the conclusion of the study tour, the Nigerian participants requested USDA/NASS assistance to pilot test an area sampling frame for use in collecting agricultural data. USDA/NASS experts traveled to Abuja, Nigeria in September 2009 to identify and evaluate resources available for developing an area

sampling frame. With input from Nigerian counterparts, it was decided to conduct the pilot in Kaduna State with the goal of having a survey in the field in May/June 2010.

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 statistically defensible methodology and produce more timely and accurate statistics. Prior to the pilot, Nigeria has used a household based Agricultural Product Survey (APS). In the APS, 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.

## II. Sampling & Estimators

### Frame Construction

Spot 5 satellite images of 5 meter resolution were used in the area frame construction, along with GIS layers for roads, railroads, rivers, political boundaries, etc. The software used was ArcGIS. Six strata were used, all based on land use. They were:

- Agricultural land, >50% cultivated (high probability of sample points finding farmers)
- Agricultural land, 15-50% cultivated (medium probability of sample points finding farmers)
- Urban/developed areas
- Agricultural land, <15% cultivated (low probability of sample points finding farmers)
- Non-agricultural
- Water (minimum 1 $km^2$)

Stratification work was done one local government area (LGA) or a contiguous group of LGAs at a time. Strata blocks were drawn off 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, the stratum code, and the land area within the block.

**Sample Selection Procedures**

Sampled points were allocated to strata as shown in table below. 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. GPS coordinates for the sample points were given to the enumerators.

The database of sampled points contained the Zone code, State code, ADP zone code, LGA code, stratum code (including total land area within stratum and the total number of points sampled within stratum), the point identification number, GPS coordinates for point, and the enumerator name and/or code. This database was used for preparing questionnaire labels, and by the edit and summary programs.

**Sample Size**

A sample of 600 points was selected. Ideally, estimates should have coefficients of variation (CVs) less than 10 percent for items of major importance. Because this was a new technique for Nigeria, it was difficult to predict what CVs will be obtained. If farm data are likely to be highly variable, the sample size to assure 10 percent 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 identify farms were maximized. The following allocation was used:

| stratum | total land (ha) | % of land | sample points | % of sample | expected expansion factor[1] |
|---|---|---|---|---|---|
| Agricultural land, >50% cultivated | 2169389.62 | 49.1% | **350** | 58.3% | 1240 |
| Agricultural land, 15-50% cultivated | 1001937.45 | 22.7% | **150** | 25.0% | 1336 |
| Urban/developed areas | 33484.03 | 0.8% | **20** | 3.3% | 335 |
| Agricultural land, <15% cultivated | 1171074.26 | 26.5% | **80** | 13.3% | 2928 |
| Non-agricultural | 36052.34 | 0.8% | | | |
| Water (minimum 1 $km^2$) | 10126.90 | 0.2% | | | |

[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, etc. The value of the population indicator was "1" if the operation satisfies 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 number of farms, the survey value was "1".

Variances, standard errors and coefficients of variation were calculated using the following formulas.

$$V(\hat{y}) = \sum_i \frac{n_i}{n_i - 1} \sum_j \left(y_{ij} - \bar{y}_i\right)^2 \text{ where } \bar{y}_i = \sum_j \frac{y_{ij}}{n_i}$$

$$SE(\hat{y}) = \sqrt{V(\hat{y})}$$

$$CV(\hat{y}) = \frac{SE(\hat{y})}{\hat{y}}$$

## III. 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). In August thru 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 October 1, 2010. All questionnaires had to be returned no later than December 15.

**Enumerator Procedures**

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 instruction on the purpose of the survey, respondent confidentiality, sampling, terms and definitions, using their 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 include a point falling on the boundary of two farms or a farm and a non-agricultural land use, a multiple family dwellings, a non-agricultural business where the operator also has a farming operation, or a farming operation has 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 one hundredth of a hectare. 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.

**Office Procedures**

The questionnaires were prepared and labeled and sent to enumerators. When questionnaires were returned to the office, they were checked-in and reviewed for completeness. If necessary, they were sent back out to the field to be completed. The data was entered in a computer database, the computer edit programs were run, and data was validated or updated as necessary. The computer edit checked for land areas and livestock items summations and basic logical relationships.

**Questionnaires**

The questionnaires contained basic land-use questions. If point fell on a farm or farmer's residence, the crops planted to field (and corresponding square meter 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-agriculture 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.

## IV. Summaries & Analysis

A computer program was developed to calculate expansions, variances, standard errors and coefficient of variations 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.

## V. 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 trained the trainers. Enumerator training was conducted in Mid-May with data collection activities beginning immediately after. 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/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, as well as 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 can 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 IT 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.