Unveiling Rwanda's Master Sampling Frame Strategy for Improving Agricultural Survey and Data Precision: A Case Study

J.C. Mwizerwa, A. Mutebutsi, A. Abayisenga, J. Gallego*

This paper reflects on the efforts of the National Institute of Statistics of Rwanda (NISR) to enhance agricultural surveys in general and the reliability of crop statistics in particular by implementing the related Master Sampling Frame (MSF) strategy, a crucial method for achieving reliable agricultural statistics and robust data systems in the country.

A MSF is a frame that can be used for several surveys, possibly in different fields. The frame developed and implemented by NISR in Rwanda currently consists of two main components: a "List Frame" (LF) of large or specialised producers, for which exhaustive data are regularly collected, and an "Area Sampling Frame" (ASF) for other agricultural holders, to ensure completeness of the crop coverage. Introduced in 2012, the ASF was initially based on segments with physical boundaries and involved a complete enumeration of the plots in the sampled segments. Subsequent methodological reviews enhanced cost-efficiency by adopting square segments and subsampling points within the sampled segments. This approach allowed NISR to reduce the number of points required per segment and to increase the number of sampled segments, leading to better coverage.

After the introduction, the paper's second section covers the historical background and the recent transformative measures implemented by the Rwandan Government to revolutionise agricultural statistics production. It highlights the importance of accurate agricultural data for evidence-based planning and decision-making, enhancing the national Gross Domestic Product (GDP), and closing data gaps for key users and stakeholders. The third section discusses the implementation process of the MSF strategy for agricultural surveys in Rwanda. It highlights the methodologies and approaches used, including the incorporation of the Geographical Information System (GIS) in MSF building and use, the evolvement of the MSF surveys, challenges encountered during the implementation of the MSF, and other surveys that help to produce additional agriculture indicators. The fourth section presents the outcomes of using the MSF for agricultural surveys, comparing its advantages over the previous systems and the improvements in data quality and precision, as well as lessons learned. Finally, the paper concludes by highlighting and summarising key takeaways from the development and implementation process, best practices for future agricultural surveys in Rwanda and beyond, and recommendations for further improvements.

Key Words: Master Sampling Frame, Agricultural Surveys, Seasonal Agricultural Survey, List Frame, Area Frame, GIS, Data precision, Policymaking, Evidence-based Planning.

* Jean Claude Mwizerwa (jicemwi@gmail.com) and Aimable Abayisenga (aimagle12@gmail.com) are respectively Economic Statistics Project Manager and Agriculture Survey Specialist, at NISR, Alexis Mutebutsi (mutebutsimushambaro@gmail.com) is Advisor to the Minister of State for Agriculture and Animal Resources at the Ministry of Agriculture and Animal Resources. Javier Gallego (jagalpi@gmail.com) is a former Professor in the University of Valladolid (Spain) and Statistician in the Joint Research Centre of the European Commission.

1. Introduction

The relevance of agricultural data in policy design for agricultural sector transformation and overall sustainable development is indisputable. Benedetti et al. (2010) reiterate that the use of comprehensive, reliable, and timely information on agricultural indicators in various practical situations arising in economic, social, and agricultural studies is more than ever recognised. Agricultural surveys are thus conducted all over the world to gather a large amount of information on the classic crops, yields, livestock, and other related agricultural resources. As a result, the statistics produced are so strongly conditioned by this vastly diverse demand that many countries, to be able to comply with these requests, strive to set up a complex system of surveys based on a harmonised and integrated set of information whose design, implementation and maintenance require a strong methodological effort.

In Rwanda, the agricultural sector represents approximately 27% of the Gross Domestic Product (GDP) (NISR, 2023). Recognising its significance to the national economy, the Government of Rwanda has built a robust National Agricultural Statistics System (NASS) founded on the Master Sampling Frame (MSF) to provide comprehensive, reliable, and accurate data, hence supporting evidence-based policy and decision-making for national planning and overall development of the agricultural sector.

This case study unveils Rwanda's endeavour to establish a robust NASS, serving as the cornerstone for all agricultural surveys in the country and, after the system was destroyed during the genocide, making it necessary to rebuild the system from scratch. The shift in efforts was initiated to notably improve the quality of crop estimates and effectively meet the rising demand for agricultural data for the sector's planning, research, and development needs.

2. Background

2.1. Historical background on agricultural statistics in Rwanda

Agriculture is the primary economic activity for most Rwandans, with 80.1% of households engaged in agricultural production on relatively small plots of land, according to the 2020 Agriculture Household Survey conducted by the National Institute of Statistics of Rwanda (NISR). The main crops include crops such as bananas, root crops such as cassava, Irish potatoes and sweet potatoes, cereals such as maize and sorghum, pulses, especially beans, and various other fruits and vegetables, and cash crops like coffee and tea. The broad types of crops have distinct growing and harvesting patterns (i.e. crop calendars[1]); this means that measurement procedures for one type of crop may not apply to another, which makes the process complex.

In the 1980s and early 1990s, Rwanda established mechanisms for gathering and analysing agricultural statistics based on standards to furnish insights into macroeconomic performance (Donovan and McKay 2004). Unfortunately, the system was destroyed during the genocide.

The Government gradually rebuilt a new agricultural data collection system. The process began with the Crop Assessment Survey (CAS), conducted by the Ministry of Agriculture and Animal Resources (MINAGRI) in collaboration with the Rwanda Agricultural Development Authority (RADA), the National Institute of Statistics of Rwanda (NISR), and the World Food Program (WFP). This survey provided essential crop estimates, which later became a key responsibility of the NISR. The assessment was solely based on a List Frame (LF) of farmers nationwide, using a sample of 2,520 households (MINAGRI, 2006). The survey provided the national crop estimates on cultivated areas, production, and yields, which were used to forecast the potential production before each season's end for food security purposes (NISR, 2012). At that time, it was the only source of regular crop statistics. Post-Harvest Agricultural Surveys have been conducted from time to time, for example, by MINAGRI from 1999 to 2001 and by NISR from 2006 to 2008. The 2008 National Agricultural Survey (NAS) was a large-scale effort, with results published in February 2010. A further source of information, albeit indirect, was the Household Integrated Living Conditions Survey (EICV[2]), an integrated survey of household living conditions in which consumption of food (both purchased and own-produced) was recorded. The first round of this survey (lasting 12 months) was conducted in 2000-2001, followed by another in 2005 to 2006, with subsequent rounds continuing every five years.

[1] *The agricultural calendar of the country is composed of the following three seasons: Season A: From September to February of the following year; Season B: From March to June; and Season C: From July to September.*

[2] *This is in French and stands for «Enquête Intégrale sur les Conditions de Vie des Ménages» (Household Integrated Living Conditions Survey).*

Analysis has been undertaken regarding EICV years (2001 and 2006, respectively) to establish the contribution of food crops to GDP. The findings revealed significant discrepancies among the three data sources: CAS, Post-Harvest Agricultural Surveys, and EICV. Broadly, but not always, the CAS data were the highest of the three in terms of food crops contribution to GDP, while the Post-Harvest Survey data tended to be the lowest. The consumption estimates from the EICV (adjusted to account for imports and exports) were somewhere between the other two sources. The discrepancies were such that they could not be solely attributed to sampling errors, indicating fundamental and systematic issues that needed to be addressed to improve the accuracy of agricultural production data. In addition, it was observed that the CAS data showed robust growth rates, which, despite reflecting several policy impacts, raised concerns that such growth might not align with the subsequent EICV findings.

Before establishing the current National Institute of Statistics of Rwanda as an autonomous organisation, the Ministry of Finance and Economic Planning (MINECOFIN) had a significant role in producing national public statistics through its Department of Statistics. The Department spearheaded a significant effort to create a new Strategic Plan for Statistics (MINECOFIN, 2002b). The NISR was created in 2005 under Organic Law No 10/2005 from the former Department of Statistics. NISR started operations practically in mid-2006 and full year operations in 2007. It was mandated to provide official statistics and coordinate the National Statistical System (Government of Rwanda, 2013).

Given the importance of the agricultural sector, a new structure was established within NISR in which Agriculture Statistics was a separate Division. Most of the staff in the division were new, while the existing staff members had limited experience in producing agricultural statistics. Therefore, building the capacity of the staff of its Agriculture Statistics Division became a high priority for NISR.

Due to emerging mutations and transformations in Rwanda's agricultural sector, and as mentioned above, the traditional Crop Assessment Survey has proven insufficient to meet users' growing data demands and update various agricultural databases, including GDP compilation (NISR, 2012). Consequently, the NISR developed a new and improved National Agricultural Statistics System capable of producing high-quality, sustainable, harmonised, and cost-effective crop production statistics for food security monitoring, program evaluation, national accounting, and other purposes. To achieve this, the NISR engaged the African Development Bank (AfDB) in 2011 to assist with assessing the current agricultural statistics system, including staffing in various institutions, and the quality and relevance of the statistics produced. The Institute thereafter hired the Agricultural Assessment International Corporation (AAIC) consultants to advise and assist the country in designing an integrated system of agriculture censuses and surveys.

Seasonal Agricultural Surveys (SAS), which are based on probability sampling and estimation methods using Multiple Frame Surveys (MFS) (ibid), were therefore recommended and introduced. The main objective of such new agricultural statistics program was to build the capacity of the NISR and the Ministry of Agriculture (MINAGRI) to produce regular, timely, accurate, credible, sustainable, harmonised, cost-effective and comprehensive agricultural statistics that would not only describe the structure of agriculture in Rwanda in terms of land use, crop production and livestock and can be used for food and agriculture policy formulation and planning, program evaluation, food security monitoring, but also which can be used for the compilation of national accounts statistics (SAS, 2013).

2.2. Significance of accurate agricultural data

Accurate agricultural data is fundamental to the effective functioning of agricultural systems, playing a pivotal role in driving progress, innovation, and sustainability in the agricultural sector. The International Food Policy Research Institute (IFPRI) states that high-quality agricultural data are essential for public policy formulation and for evaluating policy choices, and contribute to decisions on farm inputs choices, private sector investments, and donor programs, among others—which together impact agricultural and broader development outcomes (IFPRI, 2020). William et al., 2024, argued that the lack of accurate data can derail efforts in reducing food insecurity and poverty due to faulty policies informed by faulty data.

The Statistical Office of the European Commission reiterated that accurate data play a central role at all stages of the life cycle of political decision-making, from setting the stage through the preparation of decisions and setting targets to implementation monitoring and evaluation (Eurostat, 2020). In crafting effective policies, the foundation lies in timely, accurate, high-quality data and statistics. These elements provide the necessary evidence for decision-making while robust statistical systems support monitoring and evaluation processes, as emphasised by Rosero in the 2022 Statistical Yearbook of World Food and Agriculture (FAO, 2022).

2.3. Bridging agricultural data gaps: The crucial role of Master Sampling Frames

The concept of the Master Sampling Frame (MSF) for agricultural surveys is linked to its use (FAO, 2015). A sampling frame becomes a MSF if it can be used for several surveys, possibly in different fields. We focus here on MSF, which can be used for crop surveys, livestock surveys, and farm structure surveys, but it becomes more "master" if it can also be used for environmental surveys and/or forest inventories.

Multiple Frames (MF) are several sampling frames covering the same target population. Different sampling frames in a MF may overlap. Estimates of target variables are usually required in each intersection of a subset of frames, possibly leading to a relatively complex estimator. MF and the corresponding estimators have been initially studied by Hartley in the early 1960s and further developed by other authors (Hartley, 1962, Fuller and Burmeister, 1972, Lohr and Rao, 2006). A simplified situation appears when a list frame of large or specialised holdings is combined with an area frame on which holdings of the list frame can be identified during the area frame fieldwork and excluded from the data collection. In this case, the list frame can be seen as one or several strata, and the intersection of frames does not need to be considered. This situation appeared in the Rwanda MSF.

Area frames are sampling frames whose units have a geographic or spatial nature. Their units can be dimensionless points (although size can be attributed to points to account for location inaccuracy), lines of a given length, often called transects, or patches of territory, termed segments in the jargon of area frames. Area frames are well protected against bias due to incomplete coverage or double counting as long as the boundaries of the region of interest are well known. Some well-known examples of area frames are the Eurostat LUCAS survey (Eurostat, 2021), the area frame of the US National Agricultural Statistical Service (Davies, 2009) and the French TERUTI survey (Ballet, 2018). The area frame component of the Rwanda MSF is conceived for a two-stage sampling of points with square clusters of 9 ha in the first stage. Households are selected through geo-referenced sampled points with an approach previously tested by the MARS Project of the European Union (Gallego et al., 1994). Crop proportion data are collected on the sampled points, and other variables are observed by interviews with the farmer who cultivates the plot on which the sampled point has fallen. The characteristic of this area frame approach is that the area of sampled plots is not used in the formulae to estimate crop area, but it is needed to estimate other parameters, such as inputs.

The UN Statistics Commission (2010) highlighted that agricultural data in Sub-Saharan Africa is often scarce and of poor quality despite the sector's importance to so many households' livelihoods. For a long time, Rwanda's agricultural statistics depended on list frame sample surveys with area estimations based on farmers' subjectively declared data. This led to significant inconsistencies in main crop indicators, particularly in areas such as production and yield.

The MSF adopted by Rwanda for agricultural surveys has played a crucial role in bridging existing gaps by providing a comprehensive and standardised framework for sampling. This has strongly improved the guarantee that all relevant population units are included, reducing bias in data collection, analysis, and overall survey results. These estimates have gained trust and relevance by contributing to the compilation of the national GDP Food Balance Sheets used for the Ministry of Agriculture and Animal Resources planning, as well as by other users. For instance, estimates derived from seasonal agriculture surveys were utilised to set baselines for monitoring the Strategic Plan for Agriculture Transformation (PSTA-4, 2018) and remain relevant.

The integration of GIS tools in the MSF definition, sample selection and data collection has enhanced the traceability of the whole process, substantially improving the quality assurance in the computation of crop metrics such as cultivated and harvested area, production, and yield, which was achieved through precise measurement of plot areas. Improved traceability and quality assurance clearly has a positive impact on bias reduction, but estimating the possible bias is tricky and remains one of the challenges for the future. Furthermore, utilising plot areas as data collection units enables the incorporation of additional crop indicators, such as applied inputs (seeds, fertilisers, pesticides, …), across surveyed plots. It also facilitates reporting on various agricultural practices adopted by farms, such as radical terraces, irrigated lands, and agroforestry, providing comprehensive insights (NISR, 2023).

The Master Sampling Frame has laid the foundation for Rwanda to coordinate SAS better, conducted every agricultural season, with the Agricultural Household Survey (AHS), conducted every three years to capture conditions of agricultural households of the country in the context of agriculture policies and programs of the Government of Rwanda (AHS, 2018). The AHS provides additional data for various aspects related to characteristics of agricultural households, crop production, use of agricultural production, awareness of agriculture technology, the status of implementation of the government policies and programs, access to inputs, access to finance, agricultural assets, livestock numbers and other related agricultural items, using different samples (ibid).

3. Development and implementation process of the Master Sampling Frame for agricultural surveys

3.1. Design of initial agricultural surveys: Methodologies and approaches used

3.1.1. Target population and frame construction

Since 2012, Rwanda has adopted the Multiple Sampling Frame (MSF) method for agricultural surveys, which consists of combining an area sampling frame based on complete coverage of the country land and a list frame of all commercial/large farmers (in terms of acres or number of livestock) to cover crops observed rarely in all places, which are not efficiently covered by area frame approach (NISR, 2013). The MSF has been progressively improved in terms of cost-efficiency, thanks to lessons learnt in successive years (See Section 2.2. for more details). The evolution has mainly regarded the area frame component, initially based on the US-NASS area frame design, which turned out to be too expensive to develop or update in the highly fragmented landscape of Rwanda.

The area frame construction of the first survey round was constructed using high-quality satellite imagery and orthophotos from the Rwanda Natural Resource Authority (RNRA). The entire land area of Rwanda was subdivided into 10 distinct, non-overlapping homogeneous land-use strata defined according to crop intensity or other land-use characteristics (See Table 1 below). Strata were delineated based on a land cover map, with boundaries that follow physical features such as roads, paths, rivers, etc., which can be located easily on the ground. Then, a consolidated map/orthophoto constituted a land cover map, which formed a basis for area frame sampling design.

This approach was cost-efficient in the 2019 upgrade (See Section 2.2. for more details). One of the conclusions of the 2019 upgrade was that such a costly stratification does not significantly improve the efficiency of stratification compared with a much cheaper approach with strata coarsely delineated on square segments. For agricultural surveys, all the non-agricultural strata can be merged into a single excluded stratum, but the single strata need to be kept separate for possible use in environmental surveys.

Table 1
Land use stratification, 2012

Strata	Description
1	Intensive hillside cropland (50-100% cultivated)
2	Intensive marshland cropland (50-100% cultivated)
3	Extensive cropland (15-50% cultivated)
4	No-cropland (0-15% cultivated)
5	Cities & towns (0-15% cultivated)
6	Water
7	National parks (Existing boundaries given by the agency in charge of tourism)
8	Marshland, riverbeds with potential for rice (0-15%)
9	Forest
10	Tea plantation

Source: SAS, 2013

On the other hand, the list frame was constructed on large or commercial farmers based on administrative data complemented by data collected by the NISR field team on agricultural holdings identified in all districts. These holdings met specified criteria for inclusion in the survey, such as farming at least

10 hectares of land or raising a minimum number of 70 cattle and/or 350 goats, sheep, 140 pigs, 1,500 chickens, or 50 bee hives.

It is worth mentioning that Stratum 10 (tea plantations in Table 1 above) did not need to be considered in the area frame because the list frame sufficiently captures the needed information on tea producers.

3.1.2. Sampling process and data collection

Three of the ten Strata (Intensive Hillside Cropland, Intensive Marshland Cropland, and Extensive Cropland) were selected to form the area frame for agricultural survey sampling. The initial survey used a two-stage sample design to ensure national representativeness. In the first stage, each Stratum was subdivided into Primary Sampling Units (PSUs) of 100 hectares in hillside and marshland strata and 500 hectares in rangeland. PSUs within each Stratum were selected using the Probability Proportional to Size (PPS) approach and allocated proportionally within each Stratum for each District.

In the second stage, each selected PSU was further divided into Secondary Sampling Units (SSUs) or segments, approximately 20 hectares in Strata 1 and 2 or 50 hectares in Stratum 3. One SSU was then randomly selected within each PSU. In total, the survey covered 327 segments, along with a complete enumeration of large farmers (562). These farmers were identified based on specific thresholds: owning at least 10 hectares of farmland at or raising a minimum of 70 cattle, 350 goats or sheep, 140 pigs, 1,500 chicken, or 50 beehives. These thresholds were established through expert judgment, grounded in a thorough understanding of the country's commercial and intensive farming practices, as well as its livestock rearing systems, to ensure that large-scale agricultural operations were adequately represented in the survey.

All parcels were demarcated within segments using GIS tools (GPS and Map/orthophoto) and measurement materials (Measuring Tapes, Rulers, Pens, and Pencils). For large-scale farmers, Personal Digital Assistants (PDAs) were used. The data were recorded on paper questionnaires.

3.1.3. Estimation Methods

The combined estimators for a variable of interest (such as crop area, yield, production, or livestock population) incorporate data from both the area and list frames. To prevent overlap between the two frames, large farms are excluded from the area frame; however, some portions of large farms may still be present in the non-excluded strata. Since we have complete data on these large farms, the best approach is to assign zero value to grid points within these farms and adjust the number of sample point per segment, while separately including the total area and production data for large farms at the end of the estimation process.

$$\hat{Y} = \widehat{Y_A} + \widehat{Y_L}$$

- $\widehat{Y_A}$ is the estimate from the area frame after adjustment of the overlaps and $\widehat{Y_L}$ is the estimate from the list frame. The estimation is done in two steps as follows:
- $\widehat{Y_L} = \sum_{i=1}^{n} y_i$ where y_i is the variable of interest from large scale farm i, while for area frame part, a sampling weight is applied to get total estimates at stratum or national level as follows:
- $\widehat{Y_A} = \sum_{i=1}^{n} w_i y_i$ where w_i is the weight of the selected segment i and y_i is the variable of interest

3.2. Modifications introduced in 2013-2024

Since 2013, there have been various technical improvements in the SAS sample design to ensure efficiency in data collection, enhance stratification, and reduce sampling biases.

In 2014, significant improvements were made following field practical observations. These included revisiting strata definitions, which increased the number of strata from 10 to 12 to reflect better the land use homogeneity (e.g., splitting the hillside and marshland strata). The segment size was reduced from 20 to 10 hectares and 50 hectares in the rangeland stratum. Moreover, the number of sample segments increased from 327 to 540 to allow for the disaggregation of estimates at lower levels.

The new improvements in 2016/2017 were the use of point sampling to subsample plots within segments, the use of recent satellite imagery to update the land use map and the sampling frame, the revision of strata definition, and the shift from paper to Computer Assisted Personal Interview (CAPI) based data collection. These improvements significantly increased the data quality and enabled real-time monitoring and tracking of data collection progress while bringing more timely and cost-efficient method, allowing for an increase in the sample size (from 540 to 960 segments) without increasing the cost and enabling the disaggregation of survey estimates from the national/provincial level to the district level. In the standard spatial sampling wording, points are dimensionless, but size can be attributed to points as a tool to deal with location inaccuracy defining shifting rules (Eurostat, 2022). In SAS 2016/2017, a size of 10 x 10 m was attributed to each point. A random sample of 50 points was selected in each

10 hectares segment. The field enumerator used GPS devices to locate each sampled point withing segment and identified farmer managing the plot where the point was located. The enumerator then measured the plot using GIS tools and data collection applications and recording the plot's land use and crops information.

The 2016/17 round expanded the scope by introducing an Agricultural Households Survey (AHS) module to capture socio-economic characteristics and livestock information specific to agricultural households. The AHS utilised the same sampling frame as the SAS, but adopted a more comprehensive approach through an open-segment methodology. This approach collects data on the entire farm for all farms whose residences are within the segment (FAO, 1996). In addition to the three strata used in the SAS (Strata 1, 2, and 3), the AHS sampling frame introduced urban and rural strata (Strata 4.1 and 4.2), and a sample of 600 segments was drawn using a systematic sampling method. It is important to note that Strata 4.1 and 4.2 were explicitly added to create a comprehensive sampling frame for livestock estimation, as a significant portion of livestock is associated with households in villages and rural-urban areas adjacent to grazing land—areas that had previously been excluded from all previous SAS area sampling frames.

Starting from 2019/2020 SAS Season A, NISR shifted from using segments defined with physical boundaries to square segments. In addition, the number of sample points was reduced by half from 50 points to 25 points per segment. This reduction of the number of points was taken based on a simulation conducted using 2018 data.

Table 2 below presents the Coefficients of Variation (CV) for selected crops at the national level, comparing the results from the 2018 sample (50 points per cluster) with those CVs obtained by reducing the number of points while keeping the same clusters. The data shows that the increase in CV for major crops (beans and maize) is minimal when reducing the points from 50 to 5 per cluster, but the increase in CV is more important for less common crops, such as soybeans and groundnuts, and remains economically significant.

The available information was not sufficient to precisely quantify how the cost changes in function of the number of points per cluster, but Table 2 provides valuable insights. If the CV for major crops had been the only criterion for decision, 5 or 10 points per cluster would have been closer to the optimal choice, but the interest of less abundant crops suggests a more cautious reduction of the sample size. Logistic considerations and the well-being of surveyors to work in pairs supported the conservative decision to sample 25 points per cluster. Reducing the number of points per segment not only saved time and money for data collection but also allowed to increase the overall sample of segments from 960 to 1200. As a result, the survey budget has been reduced to less than half compared to the 2016/2017 design.

Table 2

Coefficients of variation for crop area estimation in 2018 (Season A)

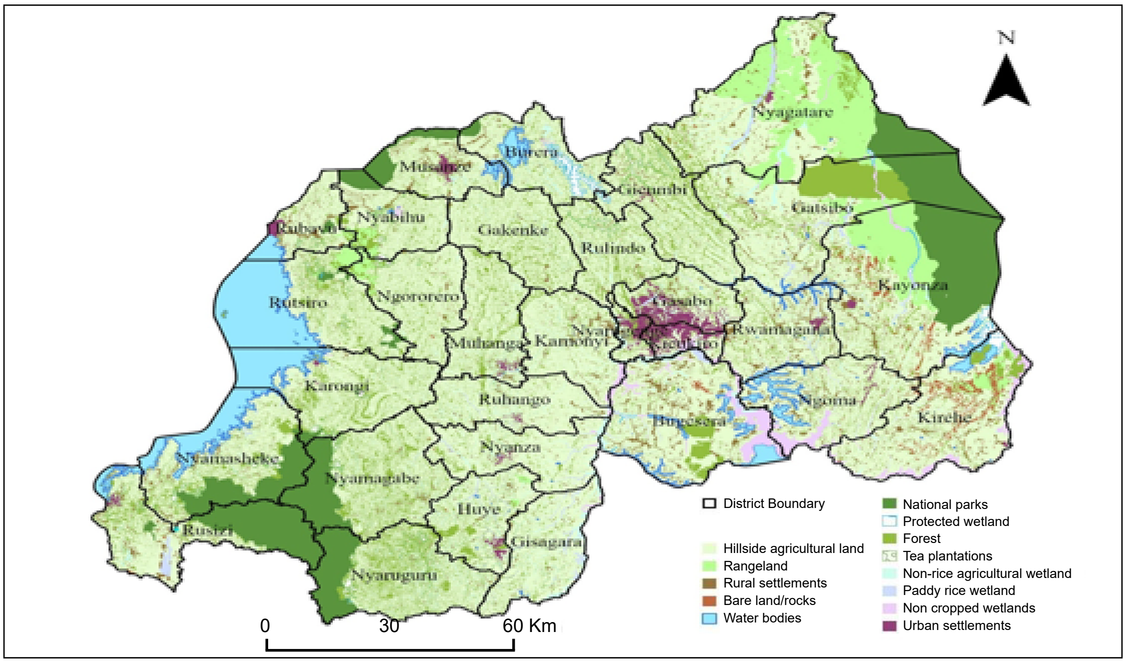
Selected Crops	Total area (1000 ha)	Points per segment			
		50	25	10	5
Beans	283	3.35%	3.37%	3.61%	3.86%
Maise	218	3.59%	3.64%	3.95%	4.40%
Soybeans	24	7.38%	8.09%	9.56%	12.80%
Ground nuts	20	7.50%	8.08%	11.01%	14.97%

Source: Gallego, 2019

Shifting from segments with physical boundaries to square segments significantly reduced the time required to update land stratification from the initial 8 months to approximately 2 months. The square segments consist of automatically dividing the country land into square or clusters, of 90,000 square meters (300m by 300m) each. Simultaneously, a land cover map was updated using recent geographical layers and high-resolution satellite imagery from Worldview (series from 2010 to 2019) (See Figure 1 below)[3]. The 300m grid is overlaid on the land cover map, and each cluster was assigned a stratum based on the dominant land cover class based on a predefined threshold as specified in Table 3 below.

[3] *This land cover data was used as a basis for constructing the area sample frame in 2019.*

Figure 1
Rwanda cover map (updated in 2019)



Source: NISR, SAS 2020

The GIS software and algorithm were developed to analyze and classify every parcel of land, resulting in a detailed land cover map for the country. In 2019, the entire land area was divided into 13 distinct land cover classes: hillside agricultural land, rangeland, rural settlements, bare land/rocks, water bodies, national parks, protected wetlands, forestry, tea plantations, non-rice marshlands, paddy rice marshlands, non-cropped wetlands, and urban settlements. This comprehensive map played a crucial role in defining the Master Sampling Frame (MSF). Furthermore, the land cover classes were further stratified into five distinct strata, based on the criteria specified in Table 3.

Table 3
Definition of strata as sets of square cells

# strat	Stratum	Definition
1	Dominant hill crops	>=60% in “hill crops” (unless >25% non-rice wetland crops)
2	Dominant Wetland crops	>25% in “wetland crops”
3	Dominant rangeland	>=60% in “rangeland” (unless >25% non-rice wetland crops).
4	Mixed clusters	Other compositions
9	Excluded clusters	>50 % in non-agricultural land cover (unless >25% wetland crops)

Source: NISR, SAS 2020

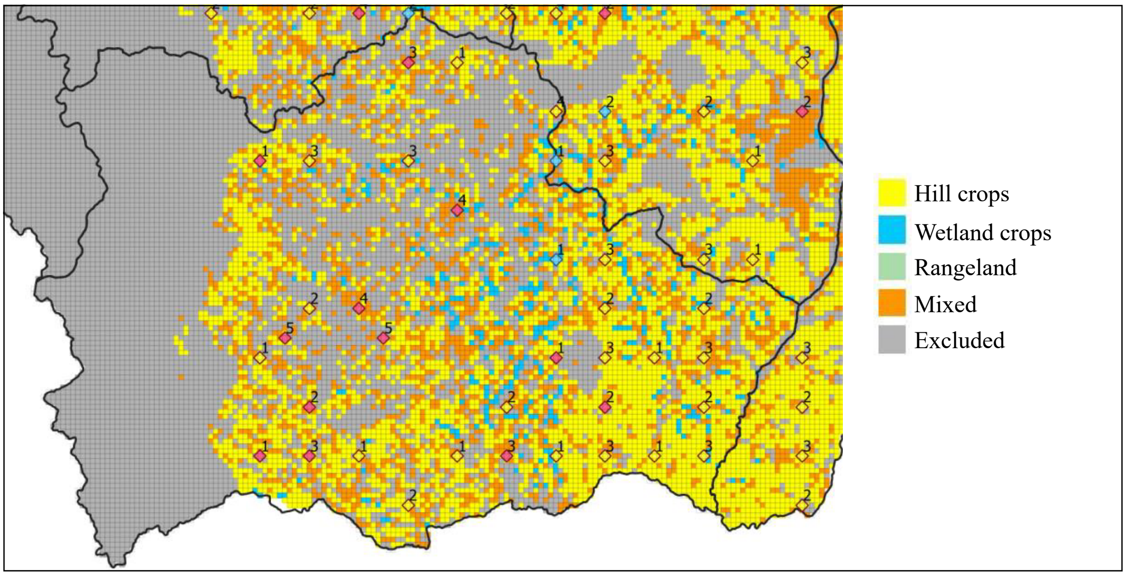
Two-dimensional stratified systematic sampling was used. Systematic sampling with a random starting point is more efficient than random sampling under general assumptions (Bellhouse, 1988). Systematic sampling also provides better traceability: an external observer can detect anomalies faster than in random sampling. In some area frame surveys, a one-dimensional systematic sampling has been used to sample PSUs that had been previously sorted with a zig-zag or serpentine approach. However, this type of systematic sampling is known to be inefficient (FAO, 2017).

A pattern of replicates is used to avoid that two clusters in the sample are too close to each other. The sampling rate was slightly increased in Stratum 2 and decreased in Stratum 3.

Figure 2 presents a district map with sample segments represented as small squares, overlaid on a color-coded background that differentiates various strata. This visualization effectively illustrates the spatial distribution of sample segments within the district, demonstrating how the approach ensures a well-balanced and representative sample coverage.

Figure 2

Stratification and sample locations in Nyaruguru[4]



Source: Gallego, 2019

For the second stage, a grid of 25 points was placed in every sampled cluster, and then, plots where those grids belong, were located using GPS, measured, and with farmers interviewed.

Reducing sample grid points per segment saves time for data collection, saves money, and allows for increasing sample segments from 960 to 1200 countrywide. Even with this increase in sample size, the annual cost of the SAS was reduced to less than half, compared with the 2016/2017 design. This cost reduction provides a major indication on relative efficiency, used to compare two different survey designs: a benchmark design 1 and a modified survey 2. The relative efficiency of design 2 compared to design 1 is:

$$rel.eff = \frac{C_1 V_1}{C_2 V_2}$$

where C_1 and V_1 are the operational cost and variance of estimates for a target variable in design 1 and C_2 and V_2 for design 2. Relative efficiency can change for different variables, so that a compromise may be necessary. For an overall concept of relative efficiency, all annual costs should be considered, including frame update and field work. Survey methodology textbooks, such as Cochran (1977), provide specific formulae that often refer to a single aspect: such as stratification, clustering or estimators that integrate some covariable. For the 2019 SAS methodological update, several aspects have changed, so that it is more practical to compute the overall efficiency with the simple formula above. Since the variances of the estimates for most variables have been similar before and after the redesign, we can conclude that the relative efficiency of the methodological update has generally been above 2. Table 2 gives a hint on the possible contribution of the cluster reduction from 50 to 25 points, but an accurate computation is not possible, because the time savings due to the cluster reduction have not been recorded.

3.3. Use of technology in frame design and data collection

Since 2012, the NISR has upgraded the agricultural survey continuously with emerging technologies. To ensure high-quality data collection of agricultural surveys and precise estimates of area data, NISR uses GPS devices with high accuracy of even less than 1 m to locate sample points and delineated plots with the support of dedicated GIS software for navigation, mapping, and analysis and processing data. Other indicators were collected, and data was recorded into tablets and directly synchronised with the servers to ensure timely data collection and monitoring. The electronic data collection application allows for quality control checks built into the data collection applications.

In recent years, with data science and machine learning development, NISR has developed an algorithm that automates some of the manual tasks involved in GIS during the land stratification process. The geo-referenced data collected from previous agricultural surveys were used to train the machine learning models used in the stratification process of SAS 2024.

[4] *The numbers displayed correspond to the replicate number.*

3.4. Challenges encountered during the implementation of the MSF

During the implementation of the area frame approach, several operational challenges emerged, affecting data quality, data collection and survey cost. Initially, the survey relied on segments with physical boundaries, which required significant labour and time to complete the whole country's land stratification. These challenges included difficulties in accurately delineating boundaries, and increased costs due to the manual effort required. However, the introduction of machine learning, automation of some tasks, and a shift from segments with physical boundaries to square segments marked a significant turning point. This innovation allowed the sample to increase from 900 to 1,200 segments while staying within the same budget. This shift not only improved the operational efficiency, but also increased the representativity and accuracy of data collected while addressing logistical and financial challenges. The area frame approach proved highly effective in producing agricultural statistics related to crop area, production, and yield. Its systematic design and spatial sampling techniques ensured robust and reliable data for these key metrics. However, it exhibited limitations in accurately estimating household-based indicators and capturing rare and special crops. Moreover, in Rwanda, where livestock is primarily raised in household systems as opposed to large-scale commercial farming, the area frame methodology was found inadequate for estimating livestock inventories. This limitation arises from the fact that livestock in Rwanda is often integrated into mixed farming systems, making it challenging to capture through area-based sampling alone. As a result, supplementary methods, such as household surveys, were deemed necessary to fill these gaps.

3.5. Other surveys that help to produce agriculture indicators

To optimise resources and based on the challenges discussed, NISR uses an integrated survey approach to capture agricultural data not covered in the regular Seasonal Agricultural Survey conducted using the Multiple Area Frame approach. Every three years, NISR conducts a Comprehensive Agricultural Survey to track and monitor household-based agricultural indicators, estimates of livestock production, health-related parameters, extension services, and more (AHS, 2020). In addition, surveys like the Labor Force Survey (LFS) are utilised to estimate agricultural labour (LSF, 2023).

In 2022, the country conducted the Population and Housing Census with an extended agricultural module and recorded geographic coordinates of all dwelling and institutional households. The extended module provided additional data on crops grown, livestock numbers, and cultivation of fruits and vegetables. The Census collected data on agricultural households (households engaged in crop cultivation and/or livestock rearing), estimated number of fruit trees, and number of livestock reared by livestock type (PHC 2022). The geographic data collected could serve as a foundation for integrating farmers' demographic and socio-economic characteristics into the Master Sampling Frame (MSF).

4. Outcomes and lessons learned

4.1. Comparative advantages with previous survey methods

A significant advantage of square segments lies in lower effort required to construct the area frame compared to the frame of segments with physical boundaries. The creation of square segments can be automated using GIS tools, a process that is both time efficient and cost effective compared to segments with physical boundaries, which rely on manual delineation based on visible physical features such as roads, rivers, or fences, which is labour-intensive and prone to inconsistencies. It is easy to update the sampling frame and sample selection when a new land cover map is available. Locating points and identifying farmers is easier for surveyors, and estimating inter-annual changes is more reliable. Over the years, with the experience gained from previous surveys and updated satellite data, the stratification process has been continuously updated and improved. This cycle of improvements has allowed for more precise and accurate classification of land strata to reflect changes in land use. For example, to improve the homogeneity within strata and better capture the diversity of land use and cropping systems, the marshland land cover class was split into three substrata: rice marshland, other crop marshland, and non-exploited marshland.

Over time, NISR has achieved significant improvements in the efficiency due to the adoption of advanced technologies, and improvement in methodologies. These efficiencies allowed to increase the sample size, ensuring the collection of more accurate and reliable estimates and data disaggregation at subnational levels such as districts and provinces.

4.2. Lessons learned

As experienced in Rwanda, the implementation of an area frame helped to update the Master Sampling

Frame every three years, significantly reducing the reliance on the typical ten years for a census cycle for this purpose. This approach not only provides more timely and relevant data, but also offers a cost-efficient alternative as the cost for acquiring satellite images and GIS tools for analysis is significantly lower compared to that of conducting a census. The updated area frame ensures comprehensive geographic coverage, reducing the risk of omitting population segments and minimising biases by accounting for structural changes in land stratification and ensuring the representation of homogeneous subgroups. In addition, combining in-person, phone, or satellite-based approaches enhances data accuracy and reliability. This method also reduces costs by focusing on specific strata or geographic areas based on land usage or the availability of other information sources. Typical examples are administrative records for tea and coffee collected by the authority overseeing their export promotion.

Regularly updating the Master Sampling Frame using latest satellite imagery and GIS tools ensures comprehensive coverage of agricultural areas by accounting for all geographic regions and changes in land use, including areas taken out of agricultural use, especially in Rwanda, where cities and urban centers are expanding. It also considers various farming practices, from minor to large-scale operations.

Leveraging modern technologies, such as remote sensing, machine learning, and mobile data collection tools enhances the accuracy and efficiency of data collection. These technologies enable remote control of data collection through dashboards, allowing for progress tracking and other controls, such as monitoring the distance between data collection points and sampled points, as well as field attendance for data quality.

The area frame also facilitates the transition to advanced data collection methods, including remote sensing, machine learning techniques and their impact on agriculture production, and GIS for precise mapping and analysis of agricultural areas.

One of the key limitations of the area frame approach is its inability to capture the livestock dimension, particularly livestock inventories, adequately. Since the area frame surveys are not designed to focus on individual households, they often miss essential household-based data, such as detailed demographic characteristics and living conditions of agricultural households, livestock information and crop data specific to households. In addition, the area frame may not adequately capture information on special crops that are not evenly spread across the country's agricultural land. To address this limitation, the National Institute of Statistics of Rwanda (NISR) conducts an Agriculture Household Survey every three years to collect detailed agricultural data at the household level to bridge this gap. The NISR collects supplementary agricultural data during the General Population and Housing Census and continuously works to enhance the administrative systems used for capturing this information.

5. Conclusion and recommendation for further improvements

Since 2012, the NISR started conducting regular Seasonal Agriculture Surveys following the three agricultural seasons of the country, namely Season A, Season B, and Season C, which helped to provide precise data on cultivated area, yield and food production regularly. This survey has significantly reduced data gaps and quality issues observed in previous surveys. Due to the flexibility and cost-effectiveness of implementing the area frame, NISR has successfully updated the area sampling frame four times in the past 12 years, in 2012, 2016-17, 2019-20, and 2023-24, respectively. These updates have increased the accuracy of collected data by regularly accounting for land use changes and monitoring the implementation of various agricultural policies. Regular updates to the Master Sampling Frame have also enabled methodological changes in data collection, incorporating less expensive and labour-intensive approaches. The regular updates to the sampling frame and the high-quality statistics generated have enhanced NISR's credibility in producing reliable data for policy development, planning and monitoring progress in the agriculture sector. It includes implementing and assessing policies such as the Strategic Plan for Agricultural Transformation.

NISR has made tremendous progress towards improving its National Agricultural Statistics System, hence the quality of produced agricultural data in general and crop data in particular. However, there is still room for further improvements, and the following are some of the required recommendations to implement in the near future, but not limited to:

- Undertaking a feasibility study on possible improvements to build a frame that can produce both area- and household-based indicators in one survey. This will help in managing resources.
- Improving the MSF applicability to specific surveys to cover livestock (in particular livestock owners without agricultural land), fishery, fruits, vegetables, and cash crop statistics.

- Use remote sensing and machine learning skills to update land cover classes and monitor crops.
- The area frame can also serve as a model in environmental surveys, as it can ensure that environmental data collection is consistent, systematic, and in line with international standards while leveraging existing infrastructure to make the process more efficient. This approach can be adapted and aligned with the Framework for the Development of Environment Statistics to provide a comprehensive system for environmental data collection in Rwanda. By expanding the area frame to include key environmental indicators, such as land use, deforestation, water resources, and biodiversity, Rwanda can enhance its capacity to monitor and respond to environmental challenges.

References

1. **Adzawla, W., Setsoafia, E. D., Setsoafia, E. D., Amoabeng-Nimako, S., Atakora, W. K., & Bindraban, P. D. (2024)**. Accuracy of agricultural data and implications for policy: Evidence from maise farmer recall surveys and crop cuts in the Guinea Savannah zone of Ghana. *Agricultural Systems*, *214*, 103817.

2. **Ballet, B., 2018**, Rénovation de l'enquête TERUTI, http://www.jms-insee.fr/2018/S23_4_ACTE_BALLET_JMS2018.pdf.

3. **Bellhouse D.R., 1988**, Systematic sampling, *Handbook of Statistics*, vol. 6, ed. P.R. Krisnaiah, C.R. Rao, pp. 125-146, North-Holland, Amsterdam.

4. **Cochran W., 1977**, *Sampling Techniques*. New York: John Wiley and Sons

5. **Davies C., 2009**, Area frame design for agricultural surveys. USDA-NASS, https://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/AREA%20FRAME%20DESIGN.pdf.

6. **Donovan, Cynthia and Andy McKay. 2004**. Comparing Agricultural Data in FSRP/DSA Surveys and the EICV Household Survey. Kigali: MINECOFIN/Department of Statistics and Oxford Policy Management.

7. **Eurostat Statistics Explained**, "LUCAS–Land use and land cover survey." URL: https://ec.europa.eu/eurostat/statistics-explained/index.php.

8. **Eurostat, 2022**, Technical reference document C1: instructions for surveyors. Section 8.4: point observation.https://ec.europa.eu/eurostat/documents/205002/13686460/C1-LUCAS-2022.pdf.

9. **FAO (2015)**, *Handbook on master sampling frames for agricultural statistics*. FAO, Rome. https://openknowledge.fao.org/.

10. **FAO 1996**, Multiple frame agricultural surveys, Volume 1, Current surveys based on area and list sampling methods.

11. **FAO, 2017**, Handbook on Remote Sensing for Agricultural Statistics. FAO, Rome, https://openknowledge.fao.org/.

12. **Gallego F.J., Delincé J., Carfagna E., 1994**, Two-Stage Area Frame Sampling on Square Segments for Farm Surveys. *Survey Methodology*, vol 20, No. 2, pp. 107-115.

13. **Fuller, W.A. & Burmeister, L.F. 1972**. Estimators for samples selected from two overlapping frames. In ASA Proceedings of the Social Statistics Section (pp. 245-249). American Statistical Association.

14. **Government of Rwanda, 2014**: Law establishing the National Institute of Statistics of Rwanda (NISR) and Determining its Mission, Organisation and Functioning.

15. **Hartley, H.O. 1962**. Multiple frame surveys. In ASA Proceedings of the Social Statistics Section (pp. 203-206). American Statistical Association.

16. **IFPRI.2020**, Political Economy of National Agricultural Statistics the Case of Niger. Discussion Paper 01929 May. https://www.ifpri.org/publication/political-economy-national-agricultural-statistics-case-niger/

17. **Javier Gallego (2019)**. Design of a new Area Frame for the Upgraded Seasonal Agricultural Survey (SAS) and the Agricultural Household Survey in Rwanda. Unpublished report.

18. **Lohr, Sharon, and JN K. Rao**. "Estimation in multiple-frame surveys." *Journal of the American Statistical Association* 101.475 (2006): 1019-1030.

19. **MINAGRI, 2018**. Strategic Plan for Agriculture Transformation 4 (PSTA 4) 2018-2024.

20. **MINECOFIN. 2002b**. Plan Stratégique De Développement du Système Statistique National 2002-2007. Kigali, Rwanda: Ministry of Finance and Economic Planning, Republic of Rwanda.

21. **National Institute of Statistics of Rwanda (NISR)**, Agricultural Household Survey 2017 report, December 2018.

22. **National Institute of Statistics of Rwanda (NISR)**, Agricultural Household Survey, December 2020.

23. **National Institute of Statistics of Rwanda (NISR)**, Labor Force Survey 2023.

24. **National Institute of Statistics of Rwanda (NISR)**, National Agricultural Survey 2008, 2010.

25. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2013, June 2015.

26. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2014, May 2015.

27. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2015, January 2016.

28. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2016, December 2016.

29. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2017, April 2018.

30. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2018, December 2018.

31. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2019, December 2019.

32. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agricultural Survey 2020, December 2020.

33. **National Institute of Statistics of Rwanda (NISR)**, Seasonal Agriculture Survey 2023.

34. **NISR, 2012**. The system of agricultural statistics in Rwanda (Improvement action plan).

35. **R. Benedetti, M. Bee, G. Espa, F. Piersimoni (2010)**. *Agricultural Survey Methods*, John Wiley & sons, DOI:10.1002/9780470665480.

36. https://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/regional_work shop_sampling_2010/FSDS_7_Multiple_frame_AS_Volume_1.pdf.

37. **UN Statistical Commission, 2017**, SDG indicators. https://unstats.un.org/sdgs/indicators/indicators-list/.

38. **William Adzawla, Edinam D. Setsoafia, Eugene D. Setsoafia, Solomon Amoabeng-Nimako, Williams K. Atakora, Prem D. Bindraban, 2024**, accuracy of agricultural data and implications for policy: Evidence from maise farmer recall surveys and crop cuts in the Guinea Savannah zone of Ghana.