



“Colombia’s experience in the development and implementation of a dual-frame sampling framework for agricultural statistics”

Viviana Andrea Guzman Rojas

National Administrative Statistics Department, Bogota D.C., Colombia.
vaguzmanr@dane.gov.co

Angela Rocio Sanabria Mosquera

National Administrative Statistics Department, Bogota D.C., Colombia.
arsanabriam@dane.gov.co

Yimer Patarroyo Quicazan

National Administrative Statistics Department, Bogota D.C., Colombia.
ypatarroyoq@dane.gov.co

Amanda Lucía Soto Agudelo

National Administrative Statistics Department, Bogota D.C., Colombia.
alsoto@dane.gov.co

Abstract

Reliable agricultural statistics are essential for evidence-based policymaking, monitoring food security, and promoting sustainable rural development. In Colombia, the long intervals between agricultural censuses and budgetary constraints resulted in the adoption of area-based sample surveys in the mid-1990s. While this approach ensured national coverage, it had limitations in terms of efficiency and precision, especially when it came to capturing large-scale agricultural producers.

Following the Third National Agricultural Census (CNA 2014), Colombia developed and implemented a dual-frame sampling framework known as the Master Rural and Agricultural Sampling Frame (MMRA). The MMRA integrates an updated area frame that covers the entire national territory and a list frame that targets large agricultural production units. This redesign enabled improved stratification and efficiency.

Domain estimation and reduced sampling errors in the National Agricultural Survey (ENA). This paper documents Colombia's methodological decisions, operational challenges, and lessons learned during the development (2018–2025) and continuous updating of the MMRA. The paper highlights how geospatial technologies, administrative records, and census data were integrated into a dynamic statistical infrastructure that aligns with international standards. The results show that a dual-frame approach can significantly strengthen official agricultural statistical systems in countries with heterogeneous agricultural structures and limited administrative coverage.

Keywords: dual-frame sampling; agricultural statistics; area frame; list frame; geospatial statistics.

1. Introduction

Historically, the production of agricultural statistics in Colombia has been shaped by financial, operational, and logistical constraints that have hindered the periodic implementation of nationwide agricultural censuses. After the second national agricultural census (CNA) in 1970, Colombia experienced an extended period without comprehensive sectoral censuses. This generated structural information gaps and necessitated alternative methodological approaches for official statistical production.

Beginning in the mid-1990s, the National Administrative Department of Statistics (DANE) adopted an area-based sampling framework as the primary strategy for producing continuous agricultural statistics. This transition was a response to budgetary restrictions and the need for periodic information to support the design and monitoring of public policies in the rural sector.

Between 1995 and 2016, DANE used an area sampling frame initially created through aerial photointerpretation and analog cartography. The rural territory was divided into relatively homogeneous geographic units, which were defined as clusters within a probabilistic sampling design. This approach was suitable given the context of highly fragmented land tenure, limited administrative records, and significant biophysical and productive heterogeneity.

Although the area frame initially ensured national territorial coverage and consistent estimation of key variables, such as land use, planted area, production, and livestock inventories, limitations emerged over time. These limitations included reduced stratification efficiency, coverage errors, limited precision in specific domains (e.g., large-scale producers), and high operational costs due to field-based updates.

The Third National Agricultural Census (CNA 2014) marked a turning point in the structure. It provided georeferenced structural information on agricultural production units (UPAs), land use, and socioeconomic characteristics. This laid the technical and conceptual groundwork for developing a dual-frame sampling approach, which was

implemented in 2017 through the Master Rural and Agricultural Sampling Frame (MMRA).

2. Conformation of the Master Rural and Agricultural Sampling Frame (MMRA) 2017.

The MMRA was constructed using structured, georeferenced information from the Third National Agricultural Census (CNA 2014), supplemented with thematic cartography and administrative records. The census provided an exhaustive inventory of agricultural (UPA) and non-agricultural (UPNA) production units, including their productive, social, and environmental characteristics, as well as their precise spatial location. The construction process began with cleaning and harmonizing the alphanumeric and geographic records and integrating the census data with the National Geo-statistical Framework (MGN). The rural territory was then subdivided into geo-statistical clusters, which are defined as continuous and mutually exclusive areas that group one or more observation units.

Cluster delimitation considered biophysical and productive homogeneity criteria, including predominant land use, thermal floor, slope range, watershed, and agroecological conditions. Incorporated secondary sources included administrative registers, cadastral information, sectoral associations, and medium- and high-resolution remote sensing products (Landsat, Sentinel, and PlanetScope).

Geospatial technology integration played a central role in this stage. The systematic use of geographic information systems (GIS) made it possible to overlay thematic layers, validate clusters cartographically, and identify spatial inconsistencies. Similarly, interpreting satellite images allowed us to confirm the classification of land use and adjust the boundaries of clusters when significant territorial changes were evident after the census operation.

A key element of the process was identifying large-scale units that contributed significantly to national agricultural production (large producers). Due to their size and productive characteristics, these units required differentiated methodological treatment. Based on census information and administrative sources, these units were organized into a list framework and subsequently integrated into the area framework. This gave rise to the dual structure of the MMRA. This approach improved statistical coverage of large producers, reduced coverage issues, and increased the accuracy of estimates.

This process resulted in the consolidation of the 2017 MMRA as a comprehensive, dynamic, georeferenced statistical framework. This framework is capable of serving as input for the design of sample surveys, the stratification of study domains, and the operational planning of field surveys simultaneously. The integration of extensive census data, updated secondary sources, and sophisticated geospatial tools paved the way for the modernization of Colombia's agricultural statistics system and facilitated the adoption of

more efficient and robust sample designs, beginning with the 2017 National Agricultural Survey (ENA).

2.1. Structure of the Dual Frame: Area Frame and List Frame

The Rural and Agricultural Master Frame (MMRA) was designed using a dual-frame approach to respond to the structural heterogeneity of the Colombian agricultural sector and overcome the limitations of previous sampling frames. This approach combines an area frame and a list frame, which are designed to capture specific segments of the target population. This maximizes statistical coverage and the efficiency of the sample design.

The area frame, which covers the entire urban and rural territory of the country, is the fundamental component of the MMRA. It consists of geostatistical clusters defined by biophysical, productive, and territorial homogeneity. These clusters group agricultural production units (UPA) and non-agricultural production units (UPNA). The MMRA ensures the inclusion of small- and medium-sized farms and dominant crops, such as cashews, avocados, lemons, carrots, and peaches, which are typically scattered in rural areas and have limited representation in administrative records, especially in mountainous or hard-to-reach regions.

The list framework was incorporated as a complementary component aimed at identifying and adequately representing large-scale agricultural production units, or large producers. These units are characterized by their high contribution to national production of goods such as coffee, rice, and palm, as well as their spatial concentration in areas with a higher level of technification. These units include agro-industrial farms and intensive livestock systems. They tend to have more structured and up-to-date administrative information, allowing them to be identified through sectoral, census, and administrative records.

Including the list frame was necessary to complement the area frame when estimating highly concentrated variables, such as planted area, production, and livestock stocks in large-scale systems. Previous designs based exclusively on areas could generate high variability between clusters, affecting the accuracy of estimates. Differentiating these farms within a list frame improved statistical efficiency and reduced sampling errors.

From a sample design perspective, adopting a dual-frame structure in the MMRA enabled the implementation of differentiated selection strategies tailored to the characteristics of different agricultural population segments. The area frame was used to probabilistically select clusters and observation units in contexts dominated by small- and medium-scale production. The list frame enabled specific schemes for large-scale units to be applied, including targeted selection or inclusion with probability one, depending on the estimation objectives' requirements. This methodological flexibility substantially improved upon the homogeneous sample designs previously used and contributed to the more efficient allocation of operational and statistical resources.

Overall, the dual-frame structure of the MMRA was a significant advance in Colombian agricultural statistics methodology. By integrating a comprehensive area frame with a specialized list frame, the MMRA balanced territorial representativeness, statistical precision, and operational feasibility. This laid the groundwork for redesigning the National Agricultural Survey sample design beginning in 2017. It also strengthened the statistical system's capacity to respond to the agricultural sector's information needs.

2.2. Definition of Geographical and Thematic Levels of the MMRA

The Rural and Agricultural Master Framework (MMRA) was structured with precise definitions of geographical and thematic levels to guarantee territorial consistency, statistical robustness, and interoperability with other official sources. This definition was based on methodological criteria specific to the sample design and the standards established in the National Geostatistical Framework (MGN). This ensures coordination between statistical and geospatial information.

2.2.1. Geographic Levels.

From a geographic perspective, the MMRA is organized hierarchically based on Colombia's Political-Administrative Division (DIVIPOLA). DIVIPOLA is the official basis for identifying departments, municipalities, population centers, and their subdivisions territorially. Using this structure, we defined clusters as the fundamental geostatistical unit of the framework within the scattered rural areas.

A cluster is a continuous territorial polygon delineated by easily identifiable natural or cultural boundaries in the field and on maps. It groups together observation units (UPA and/or UPNA). The design of the clusters took into account biophysical and productive homogeneity criteria, such as predominant land use, thermal floor, slope range, watersheds, and agroecological characteristics. This delimitation reduced internal variability and improved the efficiency of the estimators in the sample design.

Each cluster has a unique identifier structured according to DIVIPOLA coding and an internal framework numbering system. This system facilitates traceability in sample selection processes, cartographic updates, and linking with administrative records. Additionally, the MMRA incorporates supplementary geographic attributes, such as area in hectares, centroid coordinates, cluster size, hydrographic zone, slope classification, and membership in special regulatory areas, including indigenous reserves, black community lands, and national natural parks.

2.2.2. Thematic Level.

The MMRA's thematic component integrates a broad set of structural variables derived primarily from the 2014 National Agricultural Census. This set is supplemented with information from administrative records and other sectoral sources. These variables are organized into different dimensions, or components, which, together, are: Agricultural, Geographic, Livestock, Environmental, Economic, and Social. This allows for a comprehensive characterization of the observation units contained in each cluster.

Included are variables related to land use and coverage (e.g., temporary and permanent crops, forest plantations, pastures, natural forests, non-agricultural uses, and other uses), the percentage intensity of these uses within the cluster, the number of agricultural production units (APUs) and non-agricultural units (NAUs), and the property size classification.

The livestock component includes herd inventories by species and production orientation, while the economic and social dimensions consider housing, household, and people variables, as well as the presence of industrial, commercial, or service activities in the observation units. Finally, the environmental component examines the relationship between agricultural production and ecosystems. Consider aspects such as the number of UPAs that protect natural water sources, promote the sustainable use of natural resources, manage waste, and mitigate environmental impact.

3. Redefinition of Strata and Adjustment of Study Domains.

The MMRA has 100% national coverage and contains land cover variables that support the stratification criteria. The framework is updated continuously, both cartographically and thematically, as natural and administrative boundaries, land cover, and study domains change.

Although stratification generally works well, post-stratification is sometimes necessary. For example, this occurs when a significant area of crop is detected in pasture strata associated with livestock activity. Since these strata tend to have high expansion factors, an expansion factor of 1 is assigned to crop-specific variables, while the original factor is maintained for other variables in the stratum.

In areas with special regulations, some clusters are very large, so they have been divided and delimited. In regions such as the Amazon, where free images have low resolution, progress depends on institutions with higher-resolution images for support. Due to their low prevalence and territorial dispersion, crops such as beans and various fruits are difficult to represent in terms of both areas and lists.

Estimating crops is particularly challenging. In the 2019 ENA, the total area covered was 50,102,269 hectares. Of this area, 77.9% (39,017,179 hectares) was used for livestock, 10.3% (5,175,846 hectares) was forested, and 9.2% (4,617,116 hectares) was used for agriculture. With more than 200 crops having a low prevalence in relation to the total area, updating the MMRA prioritizes refining the stratification of clusters associated with agricultural use.

Redefinition of Strata and Adjustment of Study Domains.

The National Agricultural Survey is the MMRA's primary operation. It is probabilistic and stratified in design. In each stratum, clusters are selected from the area frame using simple random sampling. At the same time, the list frame is used to select farms with the largest planted area for different crops by forced inclusion. After data collection, areas reporting information from the list frame are extracted from the area frame to prevent overlap between the two frames during estimation.

The probabilistic design means that sample selection is made from the MMRA using stratification processes. In each stratum, the sample is selected using an algorithm that makes the selection probability of each element in the sampling frame known.

Stratification is based on three criteria: department, thermal floor, and predominant land use. These criteria generate a total of 341 strata. Within each stratum, clusters are selected by simple random sampling using the negative coordinate algorithm. This algorithm generates a random number for each element in a stratum and orders them in ascending order.

The list framework is based on the third CNA and is disaggregated by department and crop. It is updated annually with administrative records and information from the same surveys. Using a defined algorithm, thresholds are set for each crop to determine when a UPA is necessary for inclusion in the sample. Combining these two frames results in fewer sampling errors and greater accuracy of estimates.

4. Updating and Maintenance of the MMRA (2018–2025).

The consolidation of the MMRA in 2017 did not mark the end of its methodological development but rather the beginning of a continuous process of cartographic and thematic updates. The need for this arose from the dynamic nature of the Colombian agricultural sector, which is characterized by recurring transformations in land use and coverage, changes in productive structures, agricultural conversion processes, and variations in the location and size of production units. In this context, the sustainability and relevance of the sampling framework depended on implementing a systematic maintenance strategy that could incorporate territorial and productive changes in the rural sector in a timely manner.

Between 2018 and 2025, the MMRA underwent periodic adjustments combining cartographic, geospatial, and operational approaches. These adjustments aimed to preserve the territorial coherence of the framework, update its thematic characterization, and ensure its suitability for the changing needs of agricultural statistical operations. These adjustments reduced the risk of structural obsolescence, maintained the comparability of estimates, and strengthened the capacity of the statistical system to respond to new information demands.

MMRA update methodology: cartographic, geospatial, and operational approaches

The MMRA update methodology is structured to take the following parameters into account:

- **General Cartographic Update:** Initially, this is carried out based on updates to the departmental, municipal, and regional boundaries (populated centers and municipal capitals) of the MMRA. These updates are based on the latest version of the National Geostatistical Framework (MGN).
- **Geospatial update:** Land cover is updated based on the visual interpretation of land use in satellite images and the supervised processing of satellite images on specialized platforms, such as Google Engine.
- **Administrative Records Update:** Thematic update of the MMRA based on updated databases provided by government entities, unions, and associations in the agricultural sector, as well as lists of large producers.
- **Updating the MMRA with information from statistical operations:** This is based on the strategic integration of geographic information with data collected in specific statistical operations such as the National Agricultural Survey (ENA 2023), economic censuses, fishermen's censuses, etc.

Rather than providing a comprehensive cartographic update of the entire territory, priority is given to geographical areas (conglomerates, sample segments, etc.) that are the focus of these statistical operations. This approach optimizes resources by focusing efforts on areas of statistical interest, ensuring greater representativeness, consistency, and relevance between cartographic information and associated socioeconomic and production data. This cartographic update is carried out through field visits to clusters selected by the Directorate of Methodology and Statistical Production (DIMPE).

5. Conclusions

Challenges and limitations in the continuous updating of the MMRA.

Despite the methodological advances in updating the MMRA, ongoing maintenance faces structural, technical, financial, and operational challenges affecting its long-term sustainability. Colombia's experience demonstrates that updating the Rural and Agricultural Master Framework (MMRA) in a context of high territorial heterogeneity is a complex process requiring the balancing of technical, institutional, budgetary, and political dimensions.

The main challenges associated with updating the MMRA include the accelerated transformation of rural areas, the expansion of the agricultural frontier, the conversion of crops for production, pressure on environmentally sensitive areas, and phenomena such as illegal economies, informal mining, and armed conflicts in certain regions. These

dynamics generate spatial and productive changes that cannot always be incorporated into the sampling framework with the desired frequency.

Another significant challenge relates to the quality and scope of the administrative records used to update the listing framework. While some agricultural subsectors have consolidated information systems, the records of others contain inconsistencies, territorial gaps, and differences in classification criteria. These inconsistencies require additional cleansing, validation, and standardization processes before the information can be integrated into the MMRA. This increases the technical, economic, and operational burden of the process.

From a geospatial perspective, interpreting and processing satellite images requires specialized technical capabilities and standardized validation protocols. While tools like Google Earth can identify large-scale changes in land cover and land use, they are insufficient for territories with crop mosaics or small-scale crops. In these cases, field verification or additional sources are necessary.

Lessons Learned in the Implementation of Agricultural Dual Frames

Implementing the Rural and Agricultural Master Framework (MMRA) and integrating it into the sector's statistical operations revealed a series of relevant methodological, operational, and institutional lessons for strengthening the framework and agricultural statistics systems in contexts with high territorial heterogeneity.

One of the main lessons learned from constructing the framework initially was how to delimit Agricultural Production Units (APUs) on analog cartography. During the early stages of MMRA development, this procedure required significant time and resource investments for map scanning, manual data capture, georeferencing, polygon digitization, and subsequent cleansing and validation processes. The transition from physical media to digital environments required technical adjustments and the standardization of protocols, as well as the strengthening of institutional capacities in geospatial information management.

This process emphasized the importance of having adequate technological infrastructure and clear procedures for integrating cartographic information and statistical databases. It also demonstrated that the quality of the sampling frame largely depends on the geometric and thematic accuracy with which observation units are defined. Technological advances have substantially transformed this process in later stages. Currently, DANE has geographic applications designed for field operations that allow UPA to be delineated directly on digital cartography, supported by high-resolution satellite images. These tools facilitate real-time georeferencing, reduce errors associated with manual interpretation, and improve control over the accuracy of areas and boundaries.

Using digital cartography on mobile devices optimizes operating times, strengthens field validation mechanisms, and consolidates a structured, interoperable, updatable

geographic database. Additionally, this integration improves consistency between the information collected in the field and the geostatistical structure of the MMRA, reducing inconsistencies and facilitating subsequent spatial analysis and quality control processes.

Another important lesson is that implementing a dual sampling frame necessitates continuous coordination between DANE's departments responsible for methodological design, statistical production, and fieldwork. This coordination ensures that technical decisions in the frame's design effectively improve accuracy, coverage, and efficiency.

The Colombian experience demonstrates that the sustainability of a modern agricultural framework hinges on more than just its initial construction; it also requires the institutionalization of processes for continuous updating, technological innovation, and technical training. Transitioning to digital environments, systematically using geospatial tools, and integrating administrative records have strengthened the statistical system's ability to adapt to the rural sector's dynamics.

Taken together, these lessons demonstrate that modernizing agricultural frameworks is a progressive process combining technological innovation, methodological rigor, and institutional strengthening. The MMRA experience offers transferable elements to other national statistical offices that face similar challenges in contexts with high territorial variability and limited administrative information coverage.

References

- I. DANE. Tercer Censo Nacional Agropecuario 2014: Resultados y marco estadístico. Bogotá: Departamento Administrativo Nacional de Estadística; 2016.
- II. DANE. Modelo Genérico del Proceso Estadístico – Adaptación nacional del GSBPM. Bogotá: DANE; 2020
- III. FAO. World Programme for the Census of Agriculture 2020. Rome: Food and Agriculture Organization of the United Nations; 2017.
- IV. DANE – Dirección de Geoestadística. Metodología general actualización Marco Maestro Rural y Agropecuario MMRA. Bogotá: DANE; 2025
- V. Soto A. Experiencia de Colombia en el desarrollo del marco de muestreo dual para el sector agropecuario. Bogotá: DANE; 2025.
- VI. Stasny E. Survey sampling and frame coverage errors. *J Off Stat.* 2015;31(2):189–205