ICAS X 2026
8-10 July 2026 • Kraków, Poland

Integrating Field Surveys, Statistical Sampling, and GIS: The Random Segments Method for Agricultural Area Estimation in Argentina.

Fernando Monti

Secretaría de Agricultura, Ganadería y Pesca, Buenos Aires, Argentina-
femont@magyp.gob.ar

Maia Fiedotin

Secretaría de Agricultura, Ganadería y Pesca, Buenos Aires, Argentina-
mfiedotin@magyp.gob.ar

Abstract

Agricultural production plays a central role in Argentina's economy, contributing approximately 6% of GDP and accounting for nearly 35% of total exports through agro-industrial products. Reliable estimates of sown area are therefore essential not only for public policy and national accounts but also for producers, exporters, insurance companies, and other stakeholders.

This paper describes and evaluates the Random Segments Method (RSM), implemented by the Dirección Nacional de Estimaciones Agrícolas of Argentina for estimating the sown area of extensive crops. The RSM is a probabilistic, stratified area sampling design in which homogeneous agricultural zones are defined within each administrative unit. Randomly selected segments of approximately 400 hectares constitute the primary sampling units and are fully surveyed in the field to identify Land Use Units (LUUs). Estimates are expanded to the population level using standard stratified estimators, allowing explicit computation of variance, standard error, and coefficient of variation.

The methodology combines statistical sampling principles with geographic information systems and remote sensing tools. By maintaining fixed segments over time, the design enables temporal monitoring while preserving probabilistic properties. Currently, approximately 4,800 segments are surveyed annually across an agricultural area of nearly 50 million hectares.

Results show that the RSM provides statistically robust and transparent estimates across heterogeneous agricultural contexts. Its probabilistic nature ensures unbiased estimation and measurable precision, while field data simultaneously support crop mapping and periodic updates of the stratification. The RSM thus represents a balanced framework integrating operational feasibility, statistical rigor, and adaptability to evolving technological environments.

Keywords: Stratified area sampling; Agricultural area estimation; Sampling precision; Remote sensing; Official agricultural statistics.

[1] The authors declare that the text and materials presented in this paper are free from any copyright violations.

1. Introduction

Agricultural production in Argentina plays a strategic role in the national economy. Although primary agricultural activities account for approximately 6% of Gross Domestic Product, agro-industrial exports represent close to 35% of total national exports [1], exerting a substantial influence on trade balance, fiscal revenue, and macroeconomic stability. Beyond macroeconomic indicators, agricultural statistics are widely used by producers' organizations, exporters, input suppliers, insurance agencies, financial institutions, and research institutions.

Given this context, the accurate and timely estimation of sown area is a critical component of the official agricultural statistical system. Historically, crop area estimates relied heavily on territorial presence and expert informants within the Secretaría de Agricultura, Ganadería y Pesca (SAGyP). Field agents collected information through qualitative reports and heterogeneous survey approaches, often conditioned by local resource availability. While this system benefited from strong institutional knowledge, it lacked a fully probabilistic framework and explicit measurement of sampling error.

During the late 1990s and early 2000s, remote sensing techniques began to be incorporated into agricultural estimation. However, limited access to satellite imagery with adequate spatial and temporal resolution, persistent cloud cover, and restricted computational capacity constrained the reliability and scalability of image-based approaches. Moreover, the absence of systematic field validation reduced the inferential robustness of early implementations.

In response to these limitations and the growing demand for objective and statistically sound estimates, the Random Segments Method (RSM) was progressively developed and implemented [2]. The RSM is a stratified probabilistic area sampling design that integrates field-based observation with geographic information systems and remote sensing support. Its structure enables unbiased estimation of sown area and the explicit calculation of sampling variance, thereby ensuring statistical transparency. Currently, approximately 4,800 segments are surveyed annually across an agricultural area of nearly 50 million hectares.

Advances in satellite data availability, processing capacity, and classification algorithms have substantially expanded the analytical potential of RSM-generated data. While the RSM remains the backbone of official crop area estimation, remote sensing and GIS products now complement the design by supporting stratification updates, crop mapping, and consistency checks.

The objective of this study is to describe and evaluate the methodological foundations of the RSM implemented by SAGyP, assess its statistical properties under contrasting agricultural contexts, and discuss its integration with remote sensing-based crop mapping.

2. Methodology. The Random Segments Method

2.1 Sampling Design and Stratification

The Random Segments Method (RSM) is based on a stratified probabilistic sampling design [3,4] implemented independently within each *departamento* (administrative district within a province) conceptually aligned with area frame methodologies widely used in official agricultural statistics [5]. The target population consists of the total agricultural land within each departamento, previously delimited through remote sensing and geographic information system (GIS) procedures.

The first step in the sampling design is the definition of the sampling frame. Using satellite imagery and ancillary geographic layers, areas with null probability of agricultural activity—such as urban settlements, water bodies, rocky outcrops, and mountainous regions—are excluded from the frame. The remaining territory constitutes the agricultural domain subject to sampling.

Within this domain, the territory is partitioned into agroecologically homogeneous strata. Stratification aims to reduce within-stratum variability in cultivated area proportions and, consequently, to increase the precision of the estimators. The stratification process results in four strata:

Stratum A: areas where agriculture occupies more than 70% of the surface.

Stratum B: areas with agricultural occupation between 30% and 70%.

Stratum C: areas where agriculture covers less than 30%.

Stratum D: non-agricultural areas excluded from the sampling frame.

Historically, stratification was performed using multi-temporal NDVI classifications derived from MODIS imagery and subsequently refined with Landsat data. Currently, stratification is based on crop maps derived from multi-temporal satellite imagery using supervised remote sensing classification techniques.

To operationalize the stratification, the agricultural domain is overlaid with a hexagonal grid of approximately 2,500 hectares per cell. For each hexagon, zonal statistics are computed to determine the proportion of agricultural activity across two recent summer campaigns. Based on these proportions, each hexagon is assigned to one of the strata defined above. Subsequently, stratum boundaries are adjusted to align with significant geographic features in order to ensure spatial coherence and operational feasibility.

The rationale for stratification is threefold:

Variance reduction: by grouping areas with similar agricultural intensity, within-stratum variability decreases, improving estimator efficiency.

Statistical efficiency: stratification allows independent estimation within each stratum and weighted aggregation at the departamental level.

Comprehensive representation: the design ensures that all types of agricultural contexts—from intensive production zones to marginal areas—are proportionally represented in the sample.

Under this framework, the total cultivated area in a departamento is estimated using a stratified expansion estimator, where each stratum contributes proportionally to its total surface. The effectiveness of the design depends critically on the adequacy of the stratification to capture structural differences in agricultural intensity.

2.2 Segment Definition and Selection

Within each stratum, the primary sampling units are land segments of fixed area. Each segment covers 400 hectares and constitutes the basic unit of observation for field data collection. The use of fixed-area segments ensures uniform measurement units across strata and simplifies the expansion procedure described in Section 2.4.

Segments are selected through simple random sampling independently within each stratum. The initial step consists of generating random geographic points within the boundaries of the stratum. Each selected point determines the location of a segment according to a predefined geometric construction.

Operational Construction of Segments

Because randomly generated points frequently fall within private land without direct access, a standardized relocation algorithm is applied to ensure field operability while preserving the probabilistic nature of the design.

For each sampled point, a circular buffer is progressively expanded until it intersects a public road or accessible path. The intersection point becomes the operational anchor of the segment. From this anchor, a linear transect of 4 kilometers is defined along the road. Two parallel lines are then projected 500 meters on each side of the transect, forming a rectangular segment of 400 hectares (4 km × 1 km).

The relocation procedure follows a uniform and automated rule applied systematically to all sampled points, thereby minimizing discretion in the field and maintaining the integrity of the random selection process.

Sample Allocation Across Strata

Within each departamento, the total number of segments is fixed and remains stable over time. The allocation of segments across strata follows primarily an area-based criterion: larger strata receive a greater number of segments. However, to ensure adequate representation of all agricultural contexts, a minimum number of segments is assigned to each stratum, including marginal strata with low agricultural intensity.

This allocation scheme reflects a balance between proportional representation and coverage requirements, while contributing to the national objective of monitoring approximately 5% of the agricultural area through permanent sample segments.

Temporal Stability and Non-Response

Segments remain fixed over time, enabling longitudinal analysis and consistent interannual comparisons. This stability strengthens the robustness of time series estimates and supports cumulative database construction.

In some campaigns, operational constraints—such as temporary inaccessibility of roads due to adverse weather conditions—may prevent the survey of a small number of segments. In such cases, segments are not replaced. The estimation process relies exclusively on the effectively surveyed segments within each stratum, preserving the original probabilistic framework. The resulting reduction in effective sample size is reflected in the variance estimation and associated measures of precision.

2.3 Field Survey, Data Collection and Quality Control

Within each sampled departamento, field data collection is conducted twice per year: between March and April for summer crops, and between October and November for winter crops. During each campaign, trained field officers visit all assigned segments and directly observe land use within each segment.

Within each segment, the surface is subdivided into homogeneous land-use units (LUUs), corresponding to distinct crops or other land cover types. Each LUU is spatially delineated and assigned to a standardized and predefined list of coverage categories. Additional agronomic attributes—such as phenological stage, sowing system, and relevant field observations—may also be recorded to support classification accuracy and subsequent analysis.

Digital Data Collection Framework

Since 2015, data collection has been implemented through a dedicated mobile application that integrates preloaded segment boundaries, recent satellite imagery, geolocation functionality, and predefined categorical options. The digital system ensures spatial coherence between field observations and segment geometry, reduces transcription errors, and facilitates centralized data management. When connectivity is available, observations are synchronized with a central database.

Post-Field Validation Procedures

After field visits, collected data undergo a structured validation process. Field officers review the delineation of LUUs, verify internal consistency of coverage assignments, and perform intertemporal comparisons with historical records of the same segment. The system incorporates automated logical and spatial validation rules designed to detect inconsistencies such as overlapping polygons, incompatible land-use transitions, or implausible classifications.

Validated segments are then subject to centralized supervisory review to ensure methodological compliance and harmonization across departamentos. Only after this multi-stage validation process are data authorized for statistical expansion.

Control of Non-Sampling Error

While sampling error is formally quantified through the variance estimation described in Section 2.4, non-sampling errors—such as misclassification, measurement inaccuracies, or data processing inconsistencies—are addressed through standardized field protocols, digital data capture, automated validation rules, and supervisory controls.

This combination of probabilistic sampling and systematic quality assurance procedures contributes to the reliability and transparency of the resulting agricultural statistics.

Figure 1 shows the national coverage of the RSM and the spatial distribution of permanent segments within the agricultural domain.

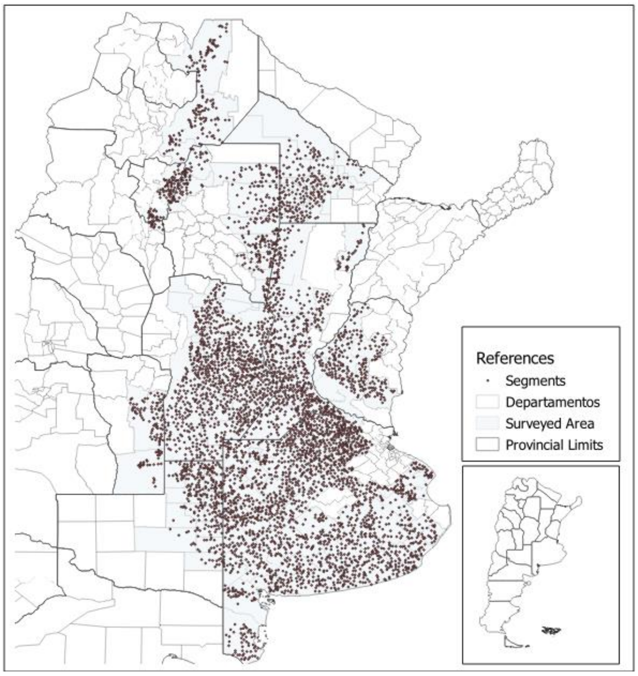

Fig. 1 *National coverage of the RSM*

2.4 Estimation and Statistical Expansion

Estimation under the RSM follows a stratified expansion approach applied independently within each departamento. Let the agricultural domain be partitioned into H strata, where A_h denotes the total area of stratum h, n_h the number of sampled segments, and y_{hi} the cultivated area observed in segment i.

The mean cultivated area per segment in stratum h is:

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$$

The estimator of total cultivated area in the departamento is:

$$\hat{Y} = \sum_{h=1}^{H} \frac{A_h}{a}\ \bar{y}_h$$

where a denotes the fixed area of each segment (400 ha). Under stratified simple random sampling within strata, this estimator is unbiased [3,4].

The variance of the estimator is computed using standard stratified sampling formulas [3,4], and the standard error is obtained as:

$$SE\left(\hat{Y}\right) = \sqrt{Var(\hat{Y})}$$

To assess relative precision, results are reported using the coefficient of variation:

$$CV\left(\hat{Y}\right) = \frac{SE(\hat{Y})}{\hat{Y}}$$

The explicit estimation of sampling error ensures statistical transparency and differentiates the RSM from non-probabilistic approaches.

Although sampling and stratification are defined at the departamento level, estimates can be aggregated to higher administrative units, such as agricultural delegations of the SAGyP. Let $\hat{Y}_d$ denote the stratified estimator for departamento d. The delegation-level estimate is obtained as:

$$\hat{Y}_D = \sum_{d\,\in\,D} \hat{Y}_d$$

where D represents the set of departamentos belonging to a given delegation.

Because segments are selected independently within each departamento, departamental-level estimators are statistically independent. Therefore, the aggregated estimator remains unbiased, and its variance is given by:

$$Var\left(\hat{Y}_D\right) = \sum_{d\,\in\,D} Var(\hat{Y}_d)$$

This additive property allows consistent precision assessment at delegation level while preserving the original stratified sampling design implemented at departamento scale.

3. Results

This section evaluates the empirical performance of the Random Segments Method for the 2024/2025 summer agricultural campaign, across four agricultural delegations: Pergamino and Marcos Juárez, representing core production regions, and Quimilí and

General Pico, representing structurally more marginal contexts. The comparison allows assessment of how sampling precision behaves under contrasting levels of agricultural intensity and crop prevalence. Figure 2 shows the location of the four SAGyP delegations included in the analysis

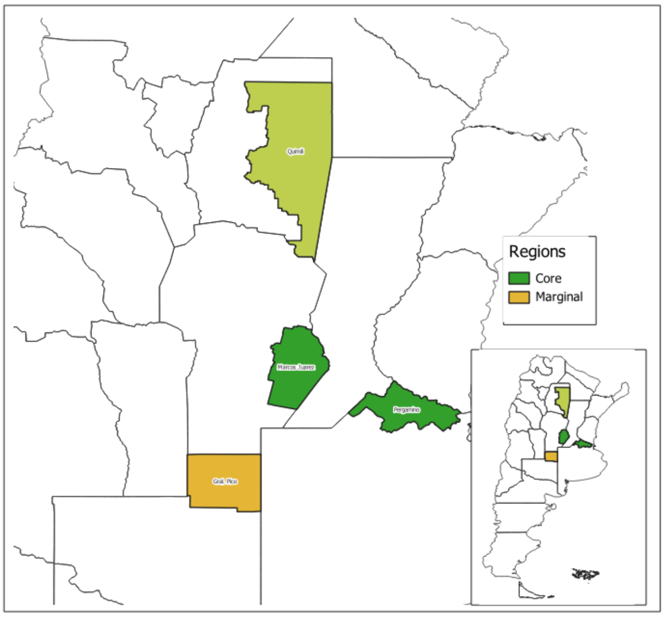

Fig. 2 *Location of the four SAGyP delegations included in the analysis*

Results are presented in terms of estimated cultivated area and coefficients of variation (CV). The CV is used as the primary indicator of relative precision. Table 1 summarizes the estimated area and coefficient of variation by crop and region.

Table 1. *Estimated sown area and coefficient of variation (CV) for selected crops and regions during the 2024/2025 summer campaign*

Region (Delegations)	Type	Crop	Estimated Area (ha)	Share of regional sown area (%)	CV (%)
Pergamino	Core	Soybean 1	382.284	37,6%	3,1%
		Soybean 2	297.821	29,3%	3,8%
		Maize 1	133.496	13,1%	7,1%
		Maize 2	32.955	3,2%	15,7%
		Sorghum 1	8.452	0,8%	21,3%
		Sorghum 2	440	0,0%	76,6%
Marcos Juárez	Core	Soybean 1	667.003	37,7%	3,1%
		Soybean 2	230.358	13,0%	7,1%
		Maize 1	407.263	23,0%	4,2%
		Maize 2	19.174	1,1%	21,1%
		Sorghum 1	28.725	1,6%	19,7%
		Sorghum 2	939	0,1%	70,2%
Quimilí	Marginal	Soybean 1	860.628	33,4%	5,4%
		Soybean 2	273.645	10,6%	8,2%
		Maize 1	358.042	13,9%	9,1%
		Maize 2	31.913	1,2%	26,8%
		Sorghum 1	43.041	1,7%	26,0%
		Sorghum 2	3.435	0,1%	46,5%
General Pico	Marginal	Soybean 1	300.394	17,8%	6,0%
		Soybean 2	31.195	1,9%	19,6%
		Maize 1	423.915	25,2%	4,4%
		Maize 2	54.262	3,2%	17,3%
		Sorghum 1	8.428	0,5%	33,3%
		Sorghum 2	1.592	0,1%	60,4%

3.1 Precision in Core Agricultural Delegations

In Pergamino and Marcos Juárez, soybean (1st season), the dominant crop in both delegations, exhibits consistently low CV values (around 3%). This indicates high precision for the principal component of cultivated area in intensive agricultural systems.

Maize (1st season) shows moderate variability, with CV values between 4% and 7%, remaining within acceptable operational thresholds. In contrast, second-season crops display higher variability. Maize (2nd season) reaches values between 15% and 21%, reflecting its lower relative prevalence.

Sorghum presents higher CV values, particularly in second-season estimates (above 70% in both delegations). However, this behavior corresponds to extremely small cultivated surfaces—well below 1% of total sown area—indicating that the elevated relative variability is driven by crop rarity rather than instability of the sampling design.

The similarity in precision patterns between the two core delegations supports the structural robustness of the design under high agricultural intensity conditions.

3.2 Precision in Marginal Agricultural Delegations

In Quimilí and General Pico, coefficients of variation are generally higher for several crops, though the pattern is not uniform across all categories.

Soybean (1st season) maintains relatively stable CV values (between 5% and 6%), only moderately higher than in core regions. This suggests that the design preserves adequate precision for dominant crops even in more heterogeneous agricultural environments.

Maize (1st season) does not exhibit a systematic deterioration in precision, with CV values comparable to those observed in core delegations. However, maize (2nd season) shows a clearer regional gradient, reaching its highest variability in Quimilí.

Second-season and low-prevalence crops consistently present higher CV values. In particular, sorghum (2nd season) exceeds 40% in both marginal delegations. As in the core regions, these high CV values are associated with very small cultivated areas, confirming that relative variability increases as crop prevalence declines under fixed sampling intensity.

3.3 Structural Patterns of Precision

Across all four delegations, two consistent patterns emerge from the empirical results:

1. Dominant crops exhibit low coefficients of variation, regardless of regional context.
2. Relative variability increases as crop prevalence decreases, independently of whether the delegation is core or marginal.

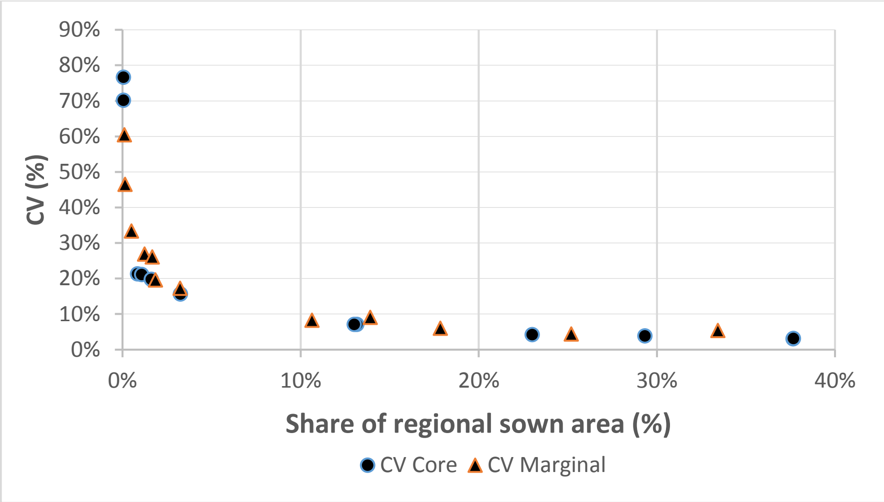

Fig. 3. *Relationship between crop share and coefficient of variation (CV) across core and marginal delegations*

Figure 3 illustrates the relationship between crop share within each regional sown area and the corresponding coefficient of variation (CV). Each point represents a crop–region combination. An inverse relationship is clearly observed: crops occupying a larger proportion of the cultivated area tend to present lower relative variability, while less prevalent crops show substantially higher CV values.

These results are consistent with theoretical expectations of stratified sampling under fixed sampling intensity. The observed increases in CV for low-prevalence crops reflect structural characteristics of agricultural distribution rather than deficiencies in the sampling design.

Overall, the RSM demonstrates stable and predictable precision behavior across heterogeneous agricultural environments, ensuring reliable estimation for crops representing the majority of cultivated area.

4. Discussion

The results confirm that the Random Segments Method (RSM) exhibits stable and predictable precision behavior across contrasting agricultural contexts. The observed patterns are consistent with theoretical expectations of stratified sampling under fixed sampling intensity, reinforcing the methodological soundness of the design.

Two structural effects emerge clearly from the empirical evaluation. First, dominant crops consistently show low coefficients of variation across both core and marginal delegations. Soybean (1st season), which represents a substantial proportion of cultivated area in all regions analyzed, maintains CV values within a narrow and operationally acceptable range. This indicates that the design effectively captures the principal components of agricultural production, even in heterogeneous environments.

Second, relative variability increases as crop prevalence decreases. Second-season crops and minor crops such as sorghum systematically present higher CV values, particularly when their share of total cultivated area is very small. This pattern is observed in both core and marginal delegations, indicating that elevated CV values are primarily driven by structural rarity rather than regional instability or weaknesses in the sampling framework.

Importantly, the comparison between core and marginal delegations reveals that regional agricultural intensity influences precision only moderately for major crops. While marginal regions exhibit slightly higher variability, the increase remains within acceptable limits for dominant crops. This suggests that the stratification strategy adequately captures differences in agricultural intensity and mitigates excessive within-stratum variability.

The behavior observed for rare crops is consistent with established sampling theory: under fixed sampling intensity, domains with very small population proportions inherently exhibit higher relative variability. In this context, high CV values for second-season sorghum do not compromise the overall reliability of the system, as these crops account for a negligible fraction of total cultivated area.

From a statistical perspective, the RSM demonstrates several strengths. The probabilistic design allows explicit quantification of sampling error, ensuring transparency and reproducibility. The stratified structure reduces variance for dominant agricultural contexts, while the fixed segment framework enables temporal stability and consistent interannual comparisons. Furthermore, the additive properties of the estimator allow aggregation across administrative levels without altering its unbiased nature.

Compared with non-probabilistic approaches based solely on remote sensing or expert judgment, the RSM provides a measurable and interpretable framework for uncertainty assessment. While remote sensing products contribute significantly to stratification and frame construction, the incorporation of probabilistic field sampling ensures that estimates are supported by formal inference and documented precision indicators, consistent with international good practices for area estimation and accuracy assessment [6,7].

Overall, the findings indicate that the RSM achieves a balance between operational feasibility and statistical rigor. The method performs robustly across heterogeneous agricultural systems, ensuring high precision for crops representing the majority of cultivated area while maintaining theoretical consistency for minor domains. These characteristics support its continued application as a reliable framework for official agricultural statistics.

5. Conclusions

This study evaluated the performance of the Random Segments Method under contrasting agricultural contexts, comparing two core production delegations and two marginal regions. The empirical results confirm that the stratified probabilistic design delivers stable and predictable precision patterns across heterogeneous agricultural systems.

Dominant crops consistently exhibit low coefficients of variation, ensuring reliable estimation for the majority of cultivated area. Increases in relative variability are primarily associated with low-prevalence crops, reflecting structural properties of the population rather than methodological deficiencies. These findings are fully consistent with theoretical expectations under fixed-intensity stratified sampling.

The ability to explicitly quantify sampling error, maintain temporal stability through permanent segments, and aggregate estimates across administrative levels without compromising unbiasedness highlights the statistical robustness of the RSM. In a context where agricultural statistics increasingly integrate remote sensing products, the incorporation of probabilistic field sampling remains essential to ensure inferential validity and transparent uncertainty measurement.

Overall, the Random Segments Method provides a balanced framework that combines operational feasibility, statistical rigor, and adaptability to diverse agricultural environments, supporting its continued use in official agricultural statistics.

References

[1] Instituto Nacional de Estadística y Censos (INDEC). Complejos exportadores, marzo de 2024 [Internet]. Informes técnicos, Vol. 8, Nº 46. Buenos Aires: INDEC; 2024 [cited 2024 Oct 1]. Available from: https://www.indec.gob.ar

[2] Secretaría de Agricultura, Ganadería y Pesca. Método de Segmentos Aleatorios Versión V. [Internet]. Buenos Aires: SAGyP; [date unknown] [cited 2025 Sept 15]. Available from: https://www.magyp.gob.ar/sitio/areas/estimaciones/metologia/

[3] Cochran WG. Sampling techniques. 3rd ed. New York: John Wiley & Sons; 1977.

[4] Lohr SL. Sampling: design and analysis. 2nd ed. Boca Raton (FL): Chapman & Hall/CRC; 2019.

[5] Food and Agriculture Organization of the United Nations. Multiple frame agricultural surveys. Volume 1: Current survey based on area and list sampling methods. Rome: FAO; 1996.

[6] Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA. Good practices for estimating area and assessing accuracy of land change. Remote Sens Environ. 2014;148:42–57.

[7] Stehman SV. Design and analysis for thematic map accuracy assessment. Remote Sens Environ. 2014;148:134–47.