Integrating Remote Sensing and Area Frame Sampling for Maize Area Estimation: Evidence from Georgia*

Anthony Burgard[1], Anna Christine Durante[1], Pamela Lapitan[1], Takaaki Masaki[†1], Arturo Y. Pacificador, Jr.[1], and Giorgi Sanadze[2]

[1]Asian Development Bank, Manila, Philippines
[2]National Statistics Office of Georgia (GEOSTAT), Agricultural and Environment Statistics Department

March 3, 2026

Abstract

Reliable and timely agricultural statistics are essential for evidence-based policymaking and food security, yet conventional field surveys are increasingly costly in fragmented farming systems. This paper tests the utility of combining area-frame probability sampling with remote sensing and machine learning for crop area estimation, using a maize application in Georgia as empirical evidence. Two independent area-frame designs were implemented in Sagarejo District: a stratified three-stage survey (1,160 points, $\sim 10 \times 10$ m) and a stratified two-stage survey (325 points). Each survey supported Sentinel-1/2–based maize classification using Random Forest and vegetation indices (NDVI, EVI), achieving overall accuracies of 95.55% and 98.60%, respectively. We then compared design-based expansion estimators to model-assisted regression estimators that used EO-derived maize maps as auxiliary information, preserving design-based inference via cross-survey training (each estimator used a map trained on the other survey). Across both designs, model-assisted estimation substantially improved precision: CV declined from 13.68% to 5.30% (61%) in the three-stage survey and from 22.25% to 4.88% (78%) in the two-stage survey. Combining the best model-assisted estimates yielded a final maize area estimate of 67.02 km^2 with CV 3.60%, a 69% gain over the combined expansion estimate. The results provide evidence that integrating probability-based area sampling with EO-driven machine learning can improve the precision and efficiency of official crop area statistics beyond a single-country setting.

Keywords: maize; agricultural statistics; remote sensing; area frame sampling; Georgia
JEL codes: C81; C83; C55; Q16

*The authors certify that the text, figures, tables, and all materials contained in this manuscript are original and free from any copyright violations.

[†]Corresponding author: tmasaki@adb.org

1. Introduction

Accurate and timely agricultural statistics form the foundation for evidence-based policymaking. Reliable crop area estimates enable consistent production forecasting, strengthen early warning systems for natural disasters and weather shocks, and support planning for agricultural inputs, subsidies, and rural infrastructure. Yet national statistics offices increasingly face resource constraints, as rising labor and transportation costs make large-scale field surveys more expensive to implement.

In Georgia, maize is a key crop for both direct consumption and livestock feed supporting the dairy and poultry sectors, which together account for up to 15 percent of agricultural gross domestic product. Despite its importance, accurate maize area estimation is challenging in a context where the average operated holding is 1.31 hectares and the average parcel size is only 0.61 hectares [1]. Land fragmentation, steep terrain, and narrow valleys increase survey costs and contribute to measurement error under conventional field-based approaches.

The growing availability of publicly accessible satellite imagery, combined with advances in cloud computing and machine learning, offers new opportunities to modernize agricultural statistics. The European Space Agency's Copernicus Programme, through the Sentinel constellation, provides free and frequent optical and radar imagery at 10 m resolution globally. When paired with robust ground reference data, these data enable the production of high-resolution crop classification maps at relatively low marginal cost. A central challenge, however, lies in integrating such maps into official statistical systems while preserving design-based inference and valid measures of uncertainty [2, 3].

An often overlooked but operationally critical issue is that the quality of machine learning–based crop maps depends fundamentally on how ground-truth data are collected. In fragmented agricultural landscapes, collecting a large number of labeled observations does not necessarily yield a representative training dataset if observations are concentrated along roads or other easily accessible areas. Such "windshield" surveys—where enumerators record land use while driving through an area—can generate high volumes of labeled points at relatively low cost. However, because roadside observations are not selected under known inclusion probabilities and tend to cluster along transport corridors, they may under-represent interior parcels and distort the class distribution of training data.

To examine this issue empirically, we implemented two independent probability-based area-frame surveys in Sagarejo District during the 2025 growing season: a stratified three-stage design with 1,160 georeferenced points and a stratified two-stage design with 325 points, yielding 1,485 ground-truth observations in total. In parallel, we conducted a windshield survey covering 1,525 roadside points (173 maize, approximately 11 percent) to benchmark how a high-volume, non-probability dataset performs relative to smaller but spatially representative area-frame samples. Sentinel-1 and Sentinel-2 imagery were processed in Google Earth Engine and classified using Random Forest algorithms to produce maize maps. To preserve statistical independence in estimation, each survey's model-assisted estimator uses an auxiliary map trained exclusively on the alternate survey's ground-truth data. The windshield dataset is used solely to generate an alternative auxiliary map and to assess how non-probability training data affect classification performance.

The results reveal two important patterns. First, maps trained on probability-based area-frame samples exhibit strong and relatively stable maize detection under independent validation, with overall accuracies above 95 percent and maize producer's and user's accuracies generally exceeding 65–80 percent. Second, the windshield-trained map—despite relying on a larger number of labeled observations—shows weaker and less consistent maize-class performance when evaluated against

probability samples, particularly in terms of maize user accuracy. These differences underscore that overall accuracy alone can mask deficiencies in minority-class detection and that spatial representativeness of training data is critical for reliable crop identification.

Beyond mapping accuracy, integrating probability-based area sampling with EO-derived auxiliary information generated substantial gains in statistical efficiency. Across both survey designs, regression estimators reduced coefficients of variation by 61–78 percent relative to conventional expansion estimators. The combined model-assisted estimate achieved a coefficient of variation of 3.60 percent, compared to 11.79 percent under combined expansion estimation. Notably, the smaller two-stage design, when paired with locally trained EO auxiliary data, achieved precision comparable to or better than the larger three-stage design under expansion alone.

Together, these findings provide two complementary insights. First, probability-based area-frame sampling remains essential for design-consistent estimation. Second, it also plays a critical role in generating high-quality training data for machine learning–based crop mapping. In fragmented farming systems, spatially balanced probability samples provide more reliable classification performance and stronger auxiliary-information gains than larger but spatially imbalanced roadside datasets.

Our contribution is threefold. First, we provide empirical evidence that integrating probability-based area-frame sampling with EO-derived auxiliary information can substantially increase statistical efficiency in official crop area estimation. Second, we demonstrate how satellite-based machine learning outputs can be formally embedded within official statistical production systems while maintaining compatibility with design-based inference. Third, we show that careful attention to sampling design in ground-truth collection is critical not only for unbiased estimation but also for ensuring robust classification performance. Together, these findings offer a scalable and statistically rigorous pathway for modernizing agricultural statistics in resource-constrained environments.

2. Study Area

Sagarejo District is located in the Kakheti region of eastern Georgia, covering 1,550.14 km^2. The district's topography ranges from 300 meters in the Iori Valley floor to approximately 1,200 meters in the Gombori mountain range. Annual precipitation averages 600–800 mm, concentrated primarily in spring (April–June) and autumn (September–November), with hot, dry summers (July–August) typical of the region's continental climate. These conditions support diverse agriculture: irrigated crops including maize and vegetables in valley bottoms, rain-fed cereals and vineyards on valley slopes, and livestock grazing on steeper hillsides.

Maize cultivation in Sagarejo follows the regional pattern: planting during April–May as soil temperatures exceed 12°C, emergence and early vegetative growth through May–June, rapid vegetative development and flowering during July–August, grain filling in August–September, and harvest during September–October. Maize is cultivated in two distinct varieties: the local variety typically grown on small-scale subsistence plots and the hybrid variety grown in larger commercial parcels used for livestock feed [4].

Administrative statistics from Sagarejo District estimate maize area between 30–45 km^2 and are in line with global estimates from the ESA WorldCereal global map (2021), which estimates 42.02 km^2 under maize, representing 2.71% of the district area.

3. Methodology

This study evaluates an integrated framework for estimating maize area that combines (i) probability-based area-frame sampling and design-based inference with (ii) Earth Observation (EO)–based crop mapping using machine learning. The framework produces two complementary outputs: (a) EO-derived maize classification maps and associated map totals used as auxiliary information, and (b) design-based maize area estimates from probability samples. We compare conventional expansion estimators to model-assisted regression (GREG) estimators and assess gains from combining independent survey designs.

To examine how training data collection methods affect both mapping accuracy and statistical estimation, we implemented two independent probability-based area-frame surveys and one non-probability windshield survey. The probability surveys provide ground truth for design-based estimation and statistically valid model-assisted inference. The windshield survey provides a high-volume but spatially imbalanced benchmark training dataset.

3.1. Sampling Design and Area Frame Construction

The sampling design was developed with two objectives: (i) provide representative ground observations for crop classification, and (ii) support design-based estimation with a target coefficient of variation (CV) below 10% under expansion estimation.

The sampling frame applied an area-frame approach using the ESA WorldCereal maize layer for Georgia (2021) as a base product for frame construction and stratification [5, 6]. The WorldCereal layer provides 10 m resolution maize classification and confidence scores [7].

Sagarejo District was overlaid with a 500 m $\times$ 500 m grid, yielding 6,201 grid cells. The grid size balances operational feasibility with within-unit homogeneity [8]. For each grid cell, we derived total 10 m pixels, maize pixels, maize proportion, and centroid coordinates (UTM Zone 38N). The frame contained 19,826,898 total pixels, of which 537,409 (2.71%) were classified as maize in WorldCereal.

3.2. Stratification

Stratification aimed to (i) reduce estimator variance, (ii) ensure representation across maize intensity gradients, and (iii) enable differential sampling rates. Grids were first separated into maize-absent (Stratum 0: zero maize pixels) and maize-present strata. Maize-present grids were then stratified using the Dalenius–Hodges cumulative root-frequency rule applied to maize pixel counts [9]. An eight-stratum configuration was selected: Stratum 0 (0 pixels), Stratum 1 (1–100), Stratum 2 (101–500), Stratum 3 (501–1,500), Stratum 4 (1,501–3,000), Stratum 5 (3,001–5,000), Stratum 6 (5,001–8,000), and Stratum 7 (>8,000).

3.3. Probability-Based Area Surveys

3.3.1 Three-Stage Design

The first probability survey used a stratified three-stage design with 1,160 observation points. Forty first-stage grids were allocated across strata ($a_0 = 2$, $a_1 = 20$, $a_2 = 6$, $a_3 = 3$, $a_4 = 3$, $a_5 = 2$, $a_6 = 2$, $a_7 = 2$), targeting an expected CV of 12.94%.

Each sampled 500 m×500 m grid was subdivided into 64 sub-grids (8×8). Eight sub-grids were selected systematically, and within each selected sub-grid five observation points (approximately 10×10 m footprint) were placed deterministically using a fixed spatial pattern. This yielded 40 observations per selected grid and 1,160 total observations.

3.3.2 Two-Stage Design

The second probability survey used a simplified stratified two-stage design with 325 observation points. Thirteen first-stage grids were selected across strata. Within each selected grid, 25 observation points were placed using a fixed spatial layout without intermediate sub-gridding.

Although smaller in size, this design provides broader spatial coverage per grid and serves as an independent probability sample for cross-validation and cross-survey training.

3.4. Windshield Survey (Non-Probability Benchmark)

To benchmark how training data collected under a non-probability approach affect crop classification performance, a windshield survey was conducted during the same fieldwork period.

Enumerators followed three preselected road corridors spanning Sagarejo District: (i) Route 1 (S5 highway, east–west, approximately 44 km), (ii) Route 2 (R38, mid-to-north, approximately 27 km), and (iii) Route 3 (R172, mid-to-south, approximately 42 km). These routes were chosen to traverse major agricultural zones and to maximize geographic coverage within logistical constraints. Teams were instructed to follow the primary routes while also entering accessible side streets when feasible to expand spatial coverage.

Data were collected by two enumerators seated in the vehicle, each observing one side of the road. Land use and crop type were recorded at approximately 100-meter intervals using GPS-enabled tablets running QField. At each observation point, the enumerator recorded the dominant land cover within the visible field parcel, along with geolocation and timestamp. This protocol yielded 1,525 labeled observations, of which 173 were maize (approximately 11 percent).

Because windshield observations were collected along accessible road networks and not selected under known inclusion probabilities, they are spatially clustered and potentially unrepresentative of the full agricultural landscape. Remote plots, fields without direct road access, and parcels located deeper within agricultural blocks are systematically less likely to be observed. For this reason, the windshield dataset is not used for design-based estimation. Instead, it is used to (i) train an alternative auxiliary crop map (X^{WS}) and (ii) assess how high-volume but spatially imbalanced training data affect classification performance and model-assisted estimation gains relative to probability-based area-frame samples.

3.5. Field Data Collection

Fieldwork was conducted June 26–28, 2025, when maize canopy development was well established. GPS-enabled tablets running Survey Solutions [10] were used for probability samples, ensuring precise geolocation. The two probability surveys yielded 1,485 ground-truth observations (453 maize; 30.5%). The windshield survey produced 1,525 roadside observations (173 maize; 11%).

The contrast in maize prevalence between probability samples (30.5%) and windshield data (11%) reflects spatial imbalance in roadside sampling and has implications for classification performance in imbalanced settings.

3.6. Satellite Data Processing and Feature Engineering

Sentinel-1 (SAR) and Sentinel-2 (optical) imagery were processed in Google Earth Engine [11]. Sentinel-2 scenes were cloud-masked using the Scene Classification Layer and aggregated into monthly median composites. Sentinel-1 imagery was radiometrically calibrated, speckle-filtered, and composited monthly. Features included vegetation indices (NDVI, EVI, RENDVI, NDWI/LSWI), brightness indices, and SAR backscatter metrics (VV, VH, VV/VH), computed per-month from median composites (Table 1). Seasonal summaries (e.g., maxima, amplitude) were optionally derived to capture phenological dynamics.

Table 1: Summary of feature set used for classification.

Category	**Features (computed per-month from monthly median composites)**
Spectral indices (S2)	NDVI = $\frac{\rho_{B8}-\rho_{B4}}{\rho_{B8}+\rho_{B4}}$; EVI = $2.5\frac{\rho_{B8}-\rho_{B4}}{\rho_{B8}+6\rho_{B4}-7.5\rho_{B2}+1}$; RENDVI = $\frac{\rho_{B5}-\rho_{B4}}{\rho_{B5}+\rho_{B4}}$; NDWI/LSWI (NIR–SWIR) = $\frac{\rho_{B8}-\rho_{B11}}{\rho_{B8}+\rho_{B11}}$.
Brightness / texture (S2)	Brightness index (BI) = $\frac{\rho_{B2}+\rho_{B3}+\rho_{B4}+\rho_{B8}}{4}$; optional GLCM texture stats (e.g., contrast/entropy) on NDVI or NIR within a moving window.
SAR metrics (S1)	Monthly medians of VV and VH backscatter (σ^0), plus VV/VH ratio (or VV_{dB}–VH_{dB}).
Phenology (S1/S2)	Monthly medians of key indices/bands (e.g., $NDVI_t$, EVI_t, VV_t, VH_t) over the growing season; optional summaries (max, min, amplitude, peak month).

Notes: ρ denotes Sentinel-2 L2A surface reflectance (bands B2, B3, B4, B5, B8, B11). Sentinel-1 features use calibrated σ^0 backscatter (VV, VH). All features are computed at 10 m; 20 m bands are resampled to 10 m in Google Earth Engine. Sentinel-2 scenes are cloud-masked using the Scene Classification Layer prior to compositing.

3.7. Classification Strategy and Independence

Random Forest classifiers [12] were trained separately using:

- Three-stage probability sample ground truth
- Two-stage probability sample ground truth
- Windshield survey ground truth

To preserve statistical independence required for valid model-assisted inference, each survey's regression estimator uses an auxiliary map trained exclusively on the other probability sample (cross-survey training). Windshield-trained maps are used only for comparative analysis and as an alternative auxiliary variable (X^{WS}).

3.8. Estimation Strategy

We estimate total maize area using:

1. Design-based expansion estimator ($\hat{Y}$),
2. Model-assisted regression estimator ($\hat{Y}_{reg}$),
3. Variance-weighted pooling of independent survey designs.

The expansion estimator is:

$$\hat{Y} = \sum_i w_i y_i,$$

where y_i is the maize indicator and w_i is the base sampling weight. Full multi-stage derivations are provided in Appendix A.1–A.2.

The model-assisted regression estimator is:

$$\hat{Y}_{reg} = \hat{Y} + b_c(X - \hat{X}),$$

where X is the auxiliary map total (e.g., X^{2S}, X^{3S}, X^{WS}, X^{WC}). Formal variance expressions are given in Appendix A.3.

Because the two probability surveys are independent, pooled estimates are computed using inverse-variance weighting:

$$\hat{Y}_{combined} = \frac{\hat{Y}_1/\hat{V}_1 + \hat{Y}_2/\hat{V}_2}{1/\hat{V}_1 + 1/\hat{V}_2}.$$

This framework allows us to directly compare how auxiliary maps trained under different ground-truth collection strategies affect statistical precision.

4. Results

4.1. Map Accuracy and Validation

Classification performance was assessed using (i) five-fold cross-validation within each training dataset and (ii) independent validation against the opposing probability sample [13]. We report overall accuracy (OA), producer's accuracy for maize (PA; sensitivity), and user's accuracy for maize (UA; precision). Results for all training–validation permutations are summarized in Table 2.

The model trained on the two-stage ground-truth sample (X^{2S}; 325 points) and validated on the three-stage sample (1,160 points) achieved an overall accuracy of 98.60%, with maize PA of 81.37% and UA of 82.14%. Conversely, the model trained on the three-stage ground-truth sample (X^{3S}; 1,160 points) and validated on the two-stage sample (325 points) achieved an overall accuracy of 95.55%, with maize PA of 67.28% and UA of 65.66%. These cross-survey results confirm strong and relatively stable maize detection when training data are derived from spatially representative probability samples.

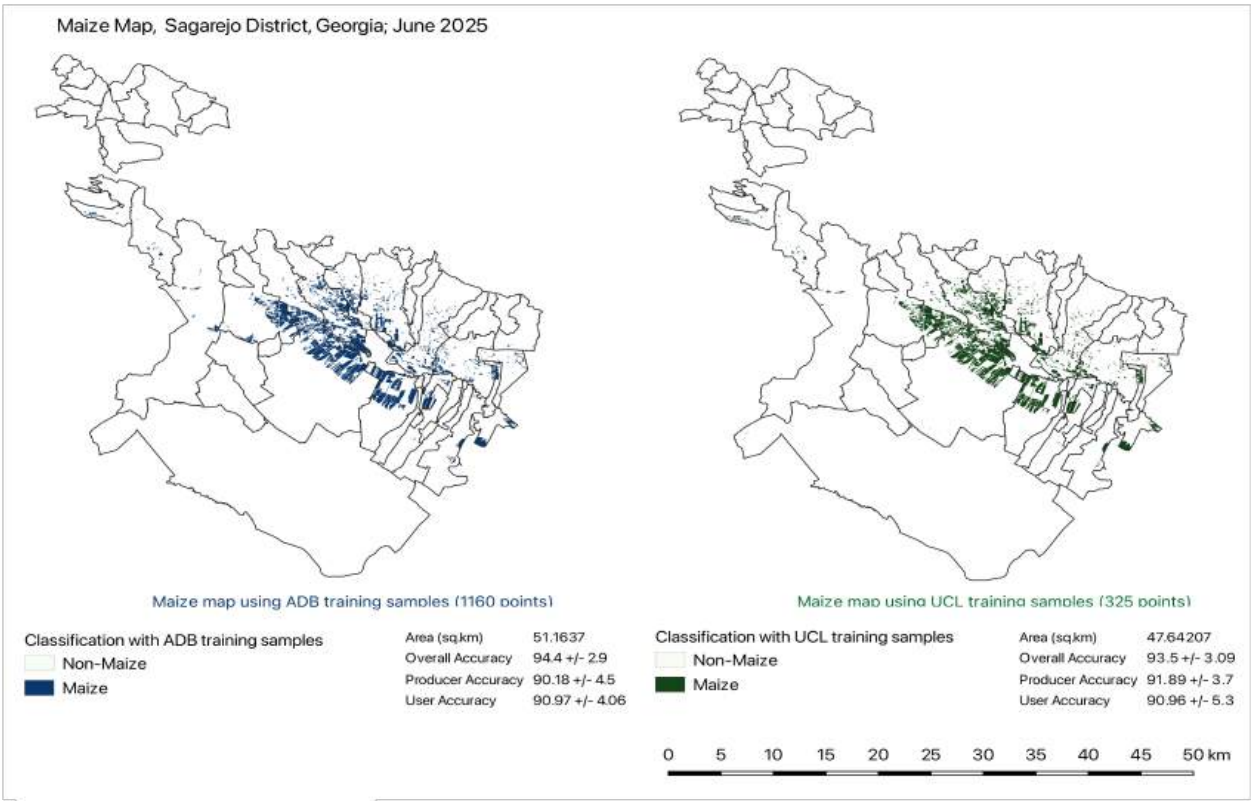
In contrast, the map trained using the windshield survey (X^{WS}), despite being based on a larger number of labeled observations (1,525 points), exhibited weaker and less consistent maize

performance under independent validation. When validated against the two-stage probability sample, the windshield-trained model achieved an overall accuracy of 93.44%, but maize PA and UA declined to 68.56% and 50.01%, respectively. Even when validated against the three-stage sample, maize UA remained substantially lower (72.10%) than the probability-sample–trained model. These results indicate reduced reliability in identifying maize parcels, particularly in terms of precision. The weaker maize performance is consistent with the spatial imbalance of the windshield dataset, which contains only 11% maize observations and is concentrated along road corridors. In class-imbalanced settings, overall accuracy can overstate classification performance because dominant non-maize classes drive the metric [13].

The relatively poorer maize-class performance of the windshield-trained model is also consistent with evidence that non-probability and convenience-based training samples can introduce spatial and class-distribution biases in land-cover classification. Because roadside observations under-represent interior and less accessible parcels, the resulting training distribution may not reflect the full variability of maize phenology and field structure across the landscape.

Applying the two probability-sample–trained models to the full district produced map-based totals of 68.73 km^2 (X^{2S}) and 64.00 km^2 (X^{3S}), both exceeding the 2021 WorldCereal baseline (42.02 km^2). These differences are consistent with both local model calibration and reported expansion of maize cultivation between 2021 and 2025. Overall, the comparison underscores that spatial representativeness of ground-truth data matters more than sample size alone for generating reliable crop maps.

Figure 1: Satellite-based maize classification maps for Sagarejo District, Georgia, June 2025 main season. Left: classification using three-stage training samples (n = 1,160); Right: classification using two-stage training samples (n = 325).



4.2. Area Frame Estimates

We now turn to design-based estimation and quantify the statistical gains from integrating EO-derived auxiliary information.

4.2.1 Design-Based Expansion Estimates

Design-based expansion estimators ($\hat{Y}$) provide the baseline against which map-assisted gains are evaluated. Using only probability-sample ground-truth data, the three-stage survey (1,160 observations; 29 grids) produced an estimate of 55.61 km^2 with CV = 13.68% (95% CI: 40.69–70.53 km^2).

The two-stage survey (325 observations; 13 grids) yielded a higher but substantially less precise estimate of 80.29 km^2 with CV = 22.25% (95% CI: 45.29–115.29 km^2).

Pooling the two independent expansion estimates using inverse-variance weights resulted in a combined estimate of 59.40 km^2 (CV = 11.79%). These results illustrate the sensitivity of pure expansion estimation to sample size and to spatial clustering of maize.

Table 2: Classification accuracy under alternative training and validation datasets. Producer's accuracy (PA) and user's accuracy (UA) are reported for maize.

Validation Data	Map (Training Data)	OA (%)	PA_{maize} (%)	UA_{maize} (%)
A. Validation using Three-Stage Probability Sample				
	X^{WC} (WorldCereal)	96.99	58.51	61.42
	X^{WS} (Windshield)	98.20	86.74	72.10
	X^{2S} (Two-stage sample)	98.60	81.37	82.14
B. Validation using Two-Stage Probability Sample				
	X^{WC} (WorldCereal)	94.02	34.77	57.33
	X^{WS} (Windshield)	93.44	68.56	50.01
	X^{3S} (Three-stage sample)	95.55	67.28	65.66

Notes: X^{3S} denotes the maize map trained using the three-stage probability sample; X^{2S} denotes the maize map trained using the two-stage probability sample; X^{WS} denotes the maize map trained using the windshield survey; X^{WC} denotes the ESA WorldCereal product. OA = overall accuracy; PA = producer's accuracy; UA = user's accuracy.

4.2.2 Model-Assisted Regression Estimates

Incorporating satellite-derived auxiliary information through regression estimation ($\hat{Y}_{reg}$) substantially improved precision across both survey designs [14, 15].

For the three-stage survey, using the EO map trained on the two-stage sample (X^{2S} = 68.73 km^2) produced a regression estimate of 62.05 km^2 with CV = 5.30% (SE = 3.29 km^2). This represents a 61% reduction in CV and approximately an 85% reduction in variance relative to the expansion estimator.

For the two-stage survey, using the EO map trained on the three-stage sample (X^{3S} = 64.00 km^2) yielded a regression estimate of 72.80 km^2 with CV = 4.88% (SE = 3.55 km^2), corresponding to a 78% reduction in CV relative to its expansion baseline.

Notably, the smaller two-stage regression estimate (325 observations; CV = 4.88%) achieved slightly higher precision than the larger three-stage regression estimate (1,160 observations; CV = 5.30%). This demonstrates that high-quality auxiliary information can compensate for smaller field samples, enabling substantial efficiency gains in official estimation.

Variance-weighted pooling of the two regression estimates yields a final combined estimate of 67.02 km^2 with CV = 3.60%, representing a 69% improvement in precision relative to the combined expansion estimate.

4.3. Role of Auxiliary Map Quality

Table 3 shows that precision gains depend critically on auxiliary map quality and local relevance. Regression using the globally trained WorldCereal product (X^{WC}) yielded only limited improvement over expansion, consistent with weaker alignment between global classification rules and local conditions. In contrast, locally calibrated EO maps (X^{2S}) generated the largest reductions in CV, reflecting stronger correlation between the auxiliary variable and the survey outcome. Regression using the windshield-trained map (X^{WS}) improved precision but less consistently, consistent with

	Area Estimate (km^2) [CV%]		
Estimator	**3-stage sample**	**2-stage sample**	**Combined**
$\hat{Y}$ (Expansion)	55.61 [13.68]	80.29 [22.25]	59.40 [11.79]
$\hat{Y}_{reg}(X^{WC})$	48.16 [12.51]	78.62 [22.54]	—
$\hat{Y}_{reg}(X^{2S})$	62.05 [5.30]	72.80 [4.88]	67.02 [3.60]
$\hat{Y}_{reg}(X^{WS})$	64.89 [6.31]	77.39 [6.31]	70.06 [4.48]

Table 3: Comparison of design-based expansion estimator ($\hat{Y}$) and model-assisted regression estimators ($\hat{Y}_{reg}$) using alternative auxiliary crop maps. X^{WC} denotes the WorldCereal global map, X^{2S} the locally trained EO map (cross-survey), and X^{WS} the windshield-trained map. The combined estimate is a variance-weighted pooling of independent survey designs.

the spatial imbalance and lower reliability of non-probability training data.

4.4. Comparison with Global Baseline Products

The final combined regression estimate of 67.02 km^2 exceeds the 2021 WorldCereal estimate of 42.02 km^2. This difference likely reflects both (i) changes in maize cultivation between 2021 and 2025 and (ii) methodological differences between globally trained products and locally calibrated models. Field observations noted increased prevalence of hybrid maize varieties consistent with rising livestock feed demand. Moreover, locally trained models better captured Sagarejo's fragmented parcels and phenological patterns than global algorithms [5, 6].

Overall, the results demonstrate that integrating probability-based area-frame sampling with locally calibrated EO auxiliary information can dramatically improve statistical efficiency, allowing smaller field samples to achieve precision levels that would otherwise require substantially larger survey designs.

5. Conclusion

This study demonstrates that integrating probability-based area-frame sampling with locally calibrated Earth Observation (EO) products can substantially enhance the precision of official crop area statistics while preserving design-based inference.

First, the results show that the representativeness of ground-truth data is more important than sample size alone. Although the windshield survey generated a larger number of labeled observations (1,525 points), maps trained on these roadside data exhibited weaker and less stable maize-class performance under independent validation, particularly in terms of maize user accuracy. By contrast, maps trained on probability-based area-frame samples achieved consistently high overall accuracies (above 95 percent) and stronger maize producer's and user's accuracies across validation datasets. These findings underscore that overall accuracy alone can mask deficiencies in minority-class detection and that spatially balanced probability samples provide more reliable training data for crop classification in fragmented agricultural systems.

Second, embedding EO-derived auxiliary information within a model-assisted regression framework generated substantial efficiency gains. Across both survey designs, coefficients of variation

declined by 61–78 percent relative to conventional expansion estimators. The final pooled regression estimate achieved a CV of 3.60 percent, representing a 69 percent precision improvement over pooled expansion. Notably, the smaller two-stage design (325 points), when paired with locally trained auxiliary data, achieved precision comparable to or better than the larger three-stage design under expansion alone.

Third, local calibration materially improved both mapping and estimation performance relative to global baseline products. Maps trained on locally collected probability samples outperformed the WorldCereal baseline in maize detection under independent validation and produced higher and more contextually consistent district-level area totals. The divergence from the 2021 WorldCereal estimate highlights the importance of updating and locally adapting crop maps for official statistical use.

Overall, the results provide a replicable and statistically rigorous pathway for modernizing agricultural statistics in resource-constrained environments. Probability-based area-frame sampling not only supports unbiased design-based estimation, but also generates high-quality training data that strengthen machine learning–based crop maps and amplify gains from model-assisted estimation. By combining probability sampling with locally calibrated EO auxiliary information, national statistical offices can achieve higher precision without proportionally increasing field sample sizes, while maintaining statistical validity and transparency.

A. Design-Based Estimation Framework

Appendix A provides the full design-based formulation of weights, estimators, and variance expressions for both survey designs, consistent with the notation used in the main text.

A.1. Expansion Estimator: Three-Stage Design (ADB)

An unbiased estimator of total maize area based on observed ground-truth data (Y) and predicted map-based values (X) is:

$$\hat{Y} = \sum_{h=1}^{L} \frac{A_h}{a_h} \sum_{i=1}^{a_h} \frac{B_{hi}}{b_{hi}} \sum_{j=1}^{b_{hi}} \frac{D_{hij}}{d_{hij}} \sum_{k=1}^{d_{hij}} y_{hijk} = \sum_{h=1}^{L} \sum_{i=1}^{a_h} \sum_{j=1}^{b_{hi}} \sum_{k=1}^{d_{hij}} w_{hijk} y_{hijk} \tag{A1}$$

$$\hat{X} = \sum_{h=1}^{L} \frac{A_h}{a_h} \sum_{i=1}^{a_h} \frac{B_{hi}}{b_{hi}} \sum_{j=1}^{b_{hi}} \frac{D_{hij}}{d_{hij}} \sum_{k=1}^{d_{hij}} x_{hijk} = \sum_{h=1}^{L} \sum_{i=1}^{a_h} \sum_{j=1}^{b_{hi}} \sum_{k=1}^{d_{hij}} w_{hijk} x_{hijk} \tag{A2}$$

The base weight is:

$$w_{hijk} = \left(\frac{A_h}{a_h}\right)\left(\frac{B_{hi}}{b_{hi}}\right)\left(\frac{D_{hij}}{d_{hij}}\right). \tag{A3}$$

Calibrated weights are:

$$w^*_{hijk} = w_{hijk} \cdot k, \quad k = \frac{A_{\text{total}}}{\sum_{h=1}^{L} \sum_{i=1}^{a_h} \sum_{j=1}^{b_{hi}} \sum_{k=1}^{d_{hij}} w_{hijk}}. \tag{A4}$$

Variance Estimation: Three-stage Design

As in the slides, variance is computed at the first-stage (grid) level:

$$s^2(\hat{Y}) = \sum_{h=1}^{L} \left(1 - \frac{a_h}{A_h}\right) \frac{a_h}{a_h - 1} \sum_{i=1}^{a_h} (y_{hi} - \bar{y}_h)^2 \tag{A5}$$

$$s^2(\hat{X}) = \sum_{h=1}^{L} \left(1 - \frac{a_h}{A_h}\right) \frac{a_h}{a_h - 1} \sum_{i=1}^{a_h} (x_{hi} - \bar{x}_h)^2 \tag{A6}$$

where

$$y_{hi} = \sum_{j=1}^{b_{hi}} \sum_{k=1}^{d_{hij}} w^*_{hijk} y_{hijk}, \quad x_{hi} = \sum_{j=1}^{b_{hi}} \sum_{k=1}^{d_{hij}} w^*_{hijk} x_{hijk},$$

and

$$\bar{y}_h = \frac{1}{a_h}\sum_{i=1}^{a_h} y_{hi}, \quad \bar{x}_h = \frac{1}{a_h}\sum_{i=1}^{a_h} x_{hi}.$$

Note: The variance estimates only takes into account the variation due to the first stage of sampling which is a common practice in multi-stage sampling. This assumes that the contribution to the sample variance of other stages are negligible.

A.2. Model-Assisted Regression Estimator: Three-stage Design

The regression estimator is:

$$\hat{Y}_{reg} = \hat{Y} + b_c\left(X - \hat{X}\right), \tag{A7}$$

where

$$b_c = \frac{\text{cov}(\hat{Y}, \hat{X})}{s^2(\hat{X})}. \tag{A8}$$

The covariance estimator is:

$$\text{cov}(\hat{Y}, \hat{X}) = \sum_{h=1}^{L}\left(1 - \frac{a_h}{A_h}\right)\frac{a_h}{a_h - 1}\sum_{i=1}^{a_h}\left(y_{hi} - \bar{y}_h\right)\left(x_{hi} - \bar{x}_h\right). \tag{A9}$$

The variance of the regression estimator is:

$$s^2(\hat{Y}_{reg}) = s^2(\hat{Y}) + b_c^2 s^2(\hat{X}) - 2b_c\,\text{cov}(\hat{Y}, \hat{X}). \tag{A10}$$

A.3. Expansion Estimator: Two-Stage Design

For the two-stage design:

$$\hat{Y} = \sum_{h=1}^{L}\sum_{i=1}^{a_h}\sum_{k=1}^{d_{hi}} w^*_{hik} y_{hik}, \quad \hat{X} = \sum_{h=1}^{L}\sum_{i=1}^{a_h}\sum_{k=1}^{d_{hi}} w^*_{hik} x_{hik}. \tag{A11}$$

The variance estimators are:

$$s^2(\hat{Y}) = \sum_{h=1}^{L}\left(1 - \frac{a_h}{A_h}\right)\frac{a_h}{a_h - 1}\sum_{i=1}^{a_h}\left(y_{hi} - \bar{y}_h\right)^2 \tag{A12}$$

$$s^2(\hat{X}) = \sum_{h=1}^{L}\left(1 - \frac{a_h}{A_h}\right)\frac{a_h}{a_h - 1}\sum_{i=1}^{a_h}\left(x_{hi} - \bar{x}_h\right)^2 \tag{A13}$$

where

$$y_{hi} = \sum_{k=1}^{d_{hi}} w^*_{hik} y_{hik}, \quad x_{hi} = \sum_{k=1}^{d_{hi}} w^*_{hik} x_{hik}.$$

References

[1] National Statistics Office of Georgia. General Information on Agricultural Holdings; 2022.

[2] Food and Agriculture Organization of the United Nations. Handbook on Remote Sensing for Agricultural Statistics. Rome: FAO; 2017.

[3] Tsiligirides TA. Remote Sensing as a Tool for Agricultural Statistics: A Review of the Area Frame Methodology with Satellite Imagery. Computers and Electronics in Agriculture. 1998;20(1):45-59.

[4] Ministry of Agriculture of Georgia. Irrigation Strategy for Georgia 2017–2025; 2017. FAOLEX document GEO171443, Food and Agriculture Organization of the United Nations. https://faolex.fao.org/docs/pdf/geo171443.pdf.

[5] Van Tricht K, et al. WorldCereal: A Dynamic Open-Source System for Global-Scale Seasonal Cropland and Crop Type Mapping at 10 m Resolution. Earth System Science Data. 2023;15:5491-518.

[6] ESA WorldCereal. WorldCereal 10 m Product Suite;. Google Earth Engine Data Catalog.

[7] Zanaga D, et al. ESA WorldCover 10 m 2021 v200. 2022.

[8] Davies C. Area Frame Design for Agricultural Surveys. Washington, DC: USDA National Agricultural Statistics Service; 2009.

[9] Dalenius T, Hodges JL. Minimum Variance Stratification. Journal of the American Statistical Association. 1959;54(285):88-101.

[10] The World Bank. Survey Solutions CAPI/CAWI platform: Release 25.01.2. Washington, DC; 2025.

[11] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. Remote Sensing of Environment. 2017;202:18-27.

[12] Breiman L. Random Forests. Machine Learning. 2001;45:5-32.

[13] Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA. Good Practices for Estimating Area and Assessing Accuracy of Land Change. Remote Sensing of Environment. 2014;148:42-57.

[14] Särndal CE, Swensson B, Wretman J. Model Assisted Survey Sampling. New York: Springer; 1992.

[15] Cassel CM, Särndal CE, Wretman J. Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. Biometrika. 1976;63(3):615-20.

[16] Burgard A, Pacificador AY Jr, Masaki T, Durante AC, Lapitan P. Mapping Rice in Viet Nam: Estimating Rice Area Using Remote Sensing and Area Frame Survey Methods in An Giang Province. Asian Development Bank; 2025.

[17] Durante AC, et al. Improving Paddy Rice Statistics Using Area Sampling Frame Technique. Asian Development Bank; 2018. 565.

[18] Deville JC, Särndal CE. Calibration Estimators in Survey Sampling. Journal of the American Statistical Association. 1992;87(418):376-82.

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

[20] Lohr SL. Sampling: Design and Analysis. 3rd ed. Boca Raton: Chapman & Hall/CRC; 2019.

[21] Stehman SV. Model-Assisted Estimation as a Unifying Framework for Estimating the Area of Land Cover and Land-Cover Change from Remote Sensing. Remote Sensing of Environment. 2009;113(11):2455-62.

[22] Foody GM. Status of Land Cover Classification Accuracy Assessment. Remote Sensing of Environment. 2002;80(1):185-201.

[23] Strahler AH, Boschetti L, Foody GM, Friedl MA, Hansen MC, Herold M, et al. Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps. Luxembourg: GOFC-GOLD; 2006. 25.

[24] Belgiu M, Drăguţ L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;114:24-31.

[25] Chen Y, Li X, Tan Z, Zeng Q. Mapping Maize Area in Heterogeneous Agricultural Landscapes Using Multi-Temporal Sentinel-1 and Sentinel-2 Data. Remote Sensing. 2021;13(15):2988.

[26] Blickensdörfer L, Schwieder M, Pflugmacher D, Hostert P. Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data. Remote Sensing of Environment. 2022;269:112831.

[27] Inglada J, Vincent A, Arias M, Marais-Sicre C. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Optical Imagery and Random Forest. Remote Sensing. 2015;7(9):12356-79.

[28] Eisfelder C, et al. Cropland and Crop Type Classification with Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine for Agricultural Monitoring in Ethiopia. Remote Sensing. 2024;16(5):866.

[29] d'Andrimont R, et al. From Parcel to Continental Scale: A First European Crop Type Map Based on Sentinel-1 and In-Situ Observations. Remote Sensing of Environment. 2021;266:112708.

[30] Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring Vegetation Systems in the Great Plains with ERTS. In: Third Earth Resources Technology Satellite-1 Symposium; 1974. p. 309-17.

[31] Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. Remote Sensing of Environment. 2002;83(1–2):195-213.

[32] Gao BC. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. Remote Sensing of Environment. 1996;58(3):257-66.

[33] Xiao X, et al. Mapping Paddy Rice Agriculture in Southern China Using Multi-Temporal MODIS Images. Remote Sensing of Environment. 2005;95(4):480-92.

[34] Xue J, Su B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of Sensors. 2017;2017:1353691.