Consolidation of the Area Framework for Agricultural Surveys

Author: Fernando Alan Martínez Hernández, National Institute of Statistics and Geography (INEGI), Aguascalientes, Mexico, alan.martinez@inegi.org

Coauthor: Susana Pérez Cadena, National Institute of Statistics and Geography (INEGI), Aguascalientes, Mexico, susana.perez@inegi.org

1. Abstract

This document[1] describes the design and challenges associated with the construction and consolidation of the Area Frame, as well as its integration with the List Frame, which together constitute the fundamental basis for future surveys related to agricultural statistics in Mexico. The main technical challenge lies in the segmentation of the national territory and the stratification of the segments according to their agricultural intensity, understood as the proportion of agricultural land relative to the total area of each segment. To achieve this, level 7 of the geo-statistical grid developed by the National Institute of Statistics and Geography (INEGI) is used, consisting of cells measuring 4,218 meters per side, covering the entire country. Segment stratification is carried out through the evaluation of various strategies that combine high-resolution satellite imagery, administrative records such as the agricultural frontier from SIAP and irrigation districts from CONAGUA, as well as the application of statistical clustering methods to group segments into five strata differentiated by their level of agricultural intensity. Additionally, polygons of urban areas and water bodies available in INEGI's National Geostatistical Framework are used to exclude areas that are not relevant to the agricultural sector. In the second stage, the stratified grid is linked to the national land parcel vector. This association allows for the connection of land parcels with production units identified in the 2022 Agricultural Census. As a result, a relationship is established between the stratified grid segments and Agricultural Production Units, and segments that could only be linked to rural land parcels are identified. This process represents a significant step forward in improving the quality and coverage of agricultural surveys in Mexico, by providing a more accurate and representative framework of the country's territory and agricultural activity. Finally, the area frame and the list frame will serve as the sampling frame for the 2025 National Agricultural Survey.

Keywords: Area frame, List frame, Agricultural Statistics, Satellite Imagery, INEGI

[1] The author declares that the text and materials presented in this manuscript are free from any copyright violations.

2. Introduction

The agricultural sector is inherently dynamic, characterized by frequent changes in crop patterns, farm structures, production technologies, and others. Farmers continuously adjust their decisions in response to market prices, climate variability, input availability, policy incentives, and access to credit. As a result, farms may change size, shift production activities, enter or exit the sector, or modify land use within short periods. Under these conditions, list frames used for agricultural surveys quickly lose accuracy and relevance.

When list frames become outdated, the planning and implementation of agricultural surveys are significantly affected. Incomplete or obsolete frames may exclude new producers, omit recently cultivated areas, or include farms that are no longer active. This can lead to under-coverage or over-coverage, introduce sampling biases, and reduce the overall quality and representativeness of survey estimates. Consequently, it becomes increasingly difficult to produce reliable statistics on planted areas, yields, and total production, particularly in regions experiencing rapid agricultural expansion or contraction.

Moreover, outdated frames hinder the statistical system's ability to detect and interpret real land-use changes. Without timely and accurate updates, it is challenging to distinguish between actual structural transformations, such as crop expansion, intensification, pasture conversion, or deforestation, and apparent variations caused by frame deficiencies.

In this regard, the Area Frame is an alternative for achieving broader coverage, since it uses the physical territory as the sampling unit, guaranteeing total and objective coverage. By dividing the map into geographic segments, it eliminates dependence on incomplete directories, reduces travel costs, and allows for direct verification on the ground, resulting in much more reliable data for economic and food-related decision-making (FAO, 2015).

According to FAO (2015), the units of an area frame can be points, transects (lines of a specific length), or pieces of land, often called segments. For area-segment frames, the main choice is between segments with physical boundaries (landscape features such as roads, rivers, or permanent field boundaries) or segments with a regular geometric shape, such as squares. The units of an area frame can be pieces of land, which are often called segments.

National Institute of Statistics and Geography (INEGI) seeks to implement the use of multiple frames, area frames and lists of statistical units, for generating agricultural statistics, with the aim of achieving broader and more accurate coverage of the primary sector in the country.

This document presents the process for constructing and consolidating the area frame, which will be implemented for the first time for the 2025 National Agricultural Survey.

3. Methodology for the consolidation of area frame

3.1. Input

To consolidate the area framework, various resources available within INEGI (National Institute of Statistics and Geography) were used, as well as information from other sources in the agricultural sector in Mexico. This was done to leverage the wealth of information and thereby simplify and reduce the time required to generate the area framework.

3.1.1 Among the most relevant inputs used from INEGI are the following:

a) 2022 Agricultural Census. This input contains all the information on agricultural production units in the country for the 2021-2022 period, as well as the land associated with them and land without agricultural activity.

b) Urban area polygons and water bodies available in the Geostatistical Framework, version corresponding to the closure of the 2022 Agricultural Census conducted by INEGI.

c) The Level 7 geostatistical grid. This grid was developed by INEGI in 2023. (Figure 1).

d) The Geostatistical Framework. It is a unique, nationwide system designed by INEGI, which presents the geostatistical division of the continental and insular territory at different levels of disaggregation This system geographically references statistical information from censuses and institutional surveys, as well as those from State Units, and integrates it into the National System of Statistical and Geographic Information (SNIEG). One of the most significant decisions was to use the institute's Level 7 geostatistical grid, which divides the country into homogeneous grids of 4,218 meters per side.

Figure 1. Division of the country into homogeneous cells of the geostatistical grid of INEGI

3.1.2 Inputs from the agricultural sector and irrigation districts of the country:

a) The Agricultural Frontier (AF) Series IV from SIAP, of the Ministry of Agriculture and Social Development. This consists of land currently under agricultural activity plus land that had activity within the last five years and is currently fallow. It is also conceptualized as the boundary zone between land occupied by crops and land not used for agricultural activities, where only natural vegetation grows, which may be used for hunting, fruit gathering, or timber extraction. The agricultural frontier was digitized at a 1:10,000 scale using visual interpretation techniques of SPOT 6 and 7 satellite imagery.

b) Polygons of the Irrigation Districts published by CONAGUA in 2024, which concentrate the most productive irrigated agricultural areas in the country.

3.2 Input linking process

Starting with the national territory segmented into uniform polygons with the Level 7 geostatistical grid, as well as a vector map of the country's delimited agricultural areas, the process consisted of intersecting these inputs to characterize each segment according to its agricultural intensity. Agricultural intensity is defined as the proportion of agricultural land within each segment.

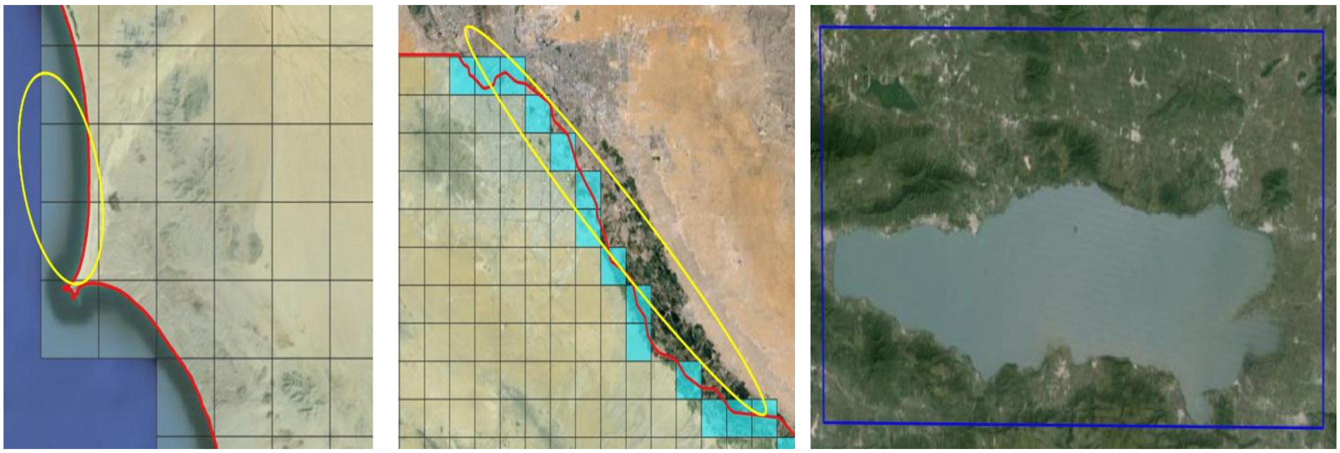
In that sense, the polygons of the irrigation districts are integrated with the polygons of the agricultural border into one. With that input at the national level, they are intersected with the grid cells to obtain the surface area and percentage of intersection, disaggregated by irrigated and rainfed areas. This makes it possible to determine the agricultural intensity of the grid cells, defined as the proportion of agricultural surface relative to the total surface area of each cell.

Additionally, the urban area and water body polygons are intersected with the grid cells to obtain surface area and percentage of intersection, in order to exclude cells that are predominantly urban. Also, the cells with less than 50% overlap with the national territory are excluded (i.e., most of the cell surface is outside the continental surface, particularly in maritime zones or in other countries) (Figure 2). In this process 1,975 grid cells were excluded.

According to the political-administrative boundaries of the National Geo-statistical Framework, state and municipality codes were assigned to the grid cells according to state with the highest percentage of intersection was assigned, followed by the municipality with the highest percentage of intersection.

Finally, the total number of segments considered within the national territory of the level 7 grid was 110, 222.

Figure 2. Exclusion of cells with 50% or more of their surface area outside the country's continental boundary, and cells in urban areas, body of water



3.3 Characterizing the Grid According to Agricultural Intensity Level.

According to FAO (2015), the stratification of an area frame is generally based on georeferenced features that can be observed on the ground, possibly through image analysis or photointerpretation. In an agricultural area frame, some typical stratum definitions might be: "agricultural land > 60 percent," "agricultural land between 30 percent and 60 percent," etc. Additional strata can be defined for specific crops or crop groups that are typically stable on the ground.

In this sense the challenge was to achieve a statistically valid stratification, from this, K-means clustering method is used to group the cells and determine the thresholds of the five strata used to characterize each grid cell, based on the percentage of intersection with the AF surface and the percentage of intersection with the irrigated surface of the AF (the latter is given greater weight because these areas are more productive).

The process was carried out at the national and state levels, as follows:

i. Five strata are predefined for grouping cells according to their agricultural intensity.

ii. The K-means method is used to statistically group cells according to their agricultural intensity into five strata (1 being the highest and 5 the lowest), using all national-level data. Two stratifications are performed: one using agricultural area variables and the other using irrigated area for the "improved FA" input.

iii. The grouping results are analyzed, and criteria are determined for each of the five strata. Irrigated areas are given greater weight in strata 1 and 2, as these contain the most productive areas in the country.

iv. The stratification results are cross-referenced and analyzed with production unit data from the 2022 Agricultural Census and with the previously performed stratification. Based on this, satellite imagery is used to review the cells with the greatest uncertainty in order to evaluate the correct classification of each one.

v. Cells are reclassified where satellite images confirm the absence of agricultural activity, or minimal activity, and which were previously classified in a high agricultural intensity stratum (or vice versa)

4. Results of Characterizing the Grid According to Agricultural Intensity Level.

The stratification criteria integrate the results of grouping both the agricultural area and irrigated land using K means, based on the variables of agricultural and irrigated agricultural proportions. Irrigated land is weighted more heavily, with lower thresholds than those for rainfed land.

The agricultural and irrigated area criteria are complementary; that is, a cell is classified into the corresponding stratum based on whether it meets the criteria for total agricultural area or those for irrigated area.

At national level, the criteria for each stratum are as follows:

Strata	Agricultural intensity	Citeria of agricultural land		Criteria of irrigated area		N° Cells
		MIN	MAX	MIN	MAX	
		>=	<	>=	<	
1	**Very high**	68.15*		48.65		5,735
2	**High**	43.9	68.13	24.96	48.65	6,557
3	**Medium**	23.5	43.9			7,440
4	**Low**	7.67	23.5			14,765
5	**Very Low**		7.67			73,750
6	**Out of sample**					1,975

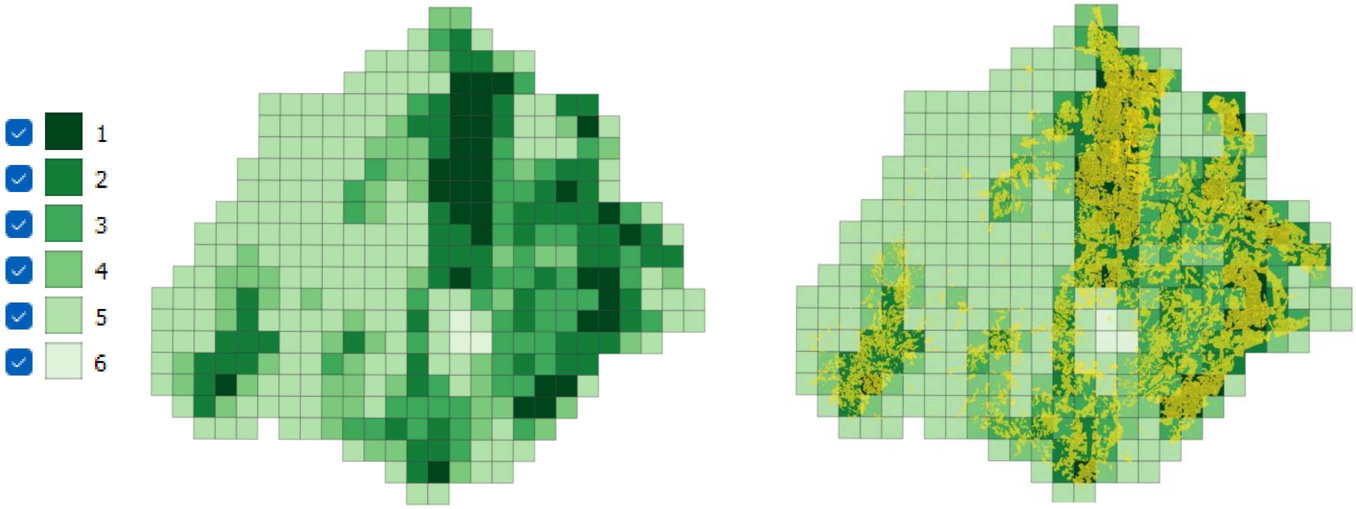
*Percentage of agricultural land within each cell

The stratification at the national and state levels is consistent with the location of agricultural areas in the country, which was reviewed using satellite images and cross-referencing various inputs, particularly with information from the 2022 agricultural census (Figure 3, 4, 5)

Figure 3. National Stratification of the cells for agricultural intensity

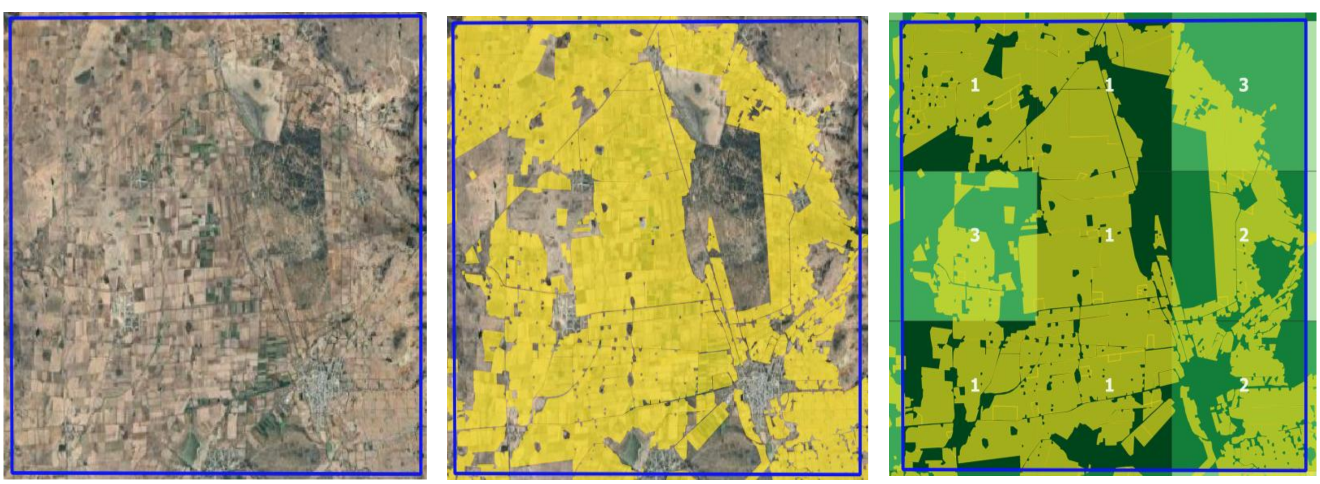
*The green colors refer to the cell classification by agricultural intensity, with dark green indicating higher intensity and lighter green indicating lower intensity.

Figure 4. Stratification of the cells of the State of Aguascalientes



* The polygons in the agricultural zone are shown in yellow.
** The green colors refer to the cell classification by agricultural intensity, with dark green indicating higher intensity and lighter green indicating lower intensity.

Figure 5. Rural area in the country categorized by its agricultural intensity



* The polygons in the agricultural zone are shown in yellow.
** The green colors refer to the cell classification by agricultural intensity, with dark green indicating higher intensity and lighter green indicating lower intensity.

6. Conclusions

The characterization of the Level 7 geo-statistical grid according to agricultural intensity provides a consistent and spatially robust framework to consolidate the area frame for agricultural surveys.

By integrating the Agricultural Frontier, urban and water body polygons, and a clustering-based stratification methodology, the approach establishes objective thresholds to differentiate areas according to their degree of agricultural presence and productivity.

The use of K-means clustering allows for a data-driven definition of strata, while the explicit consideration of irrigated areas, due to their higher productive capacity, strengthens the analytical relevance of the classification.

The identification of an out of sample stratum ensures methodological consistency by excluding predominantly urban and non-potential new agricultural areas from the area frame.

The statistical analysis reveals that while k-means clustering show high volatility in Stratum 5, regionalization serves as a more effective lens than state-level stratification for stabilizing the coefficient of variation.

Methodological refinement can increase robustness and adaptability adding remote-sensing layers would make strata responsive to land use change. Validation with independent field data or high-resolution imagery could be incorporated to quantify accuracy and provide elements for restratifying the segments

Overall, this methodology improves the statistical infrastructure to consolidate the area framework for agricultural surveys.

5. References

I. CONAGUA, Sistema Nacional de Información del Agua, [Internet] 2024, Available from: https://sinav30.conagua.gob.mx:8080/

II. FAO. Manual sobre Marcos Maestros de Muestreo para las Estadísticas Agropecuarias [Internet]. Rome: Food and Agriculture Organization of the United Nations; 2015 [cited 2026 Mar 2]. Available from: https://openknowledge.fao.org/server/api/core/bitstreams/81ecdd17-3858-48c4-83ae-1c369e140576/content

III. National Institute of Statistics and Geography (Mexico). Agricultural Census 2022 (Censo Agropecuario 2022). Mexico City: INEGI; 2022.

IV. National Institute of Statistics and Geography (Mexico). Metodología de la malla para la publicación de información estadística y geográfica) [Internet]. Mexico City: INEGI; 2023 Available from: https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/889463912002.pdf

V. SIAP. Frontera Agrícola [Internet]. 2022 Available from: https://www.gob.mx/agricultura/articulos/sabes-que-es-la-frontera-agricola