

AREA SAMPLING AND INFORMATION SYSTEMS APPLIED TO LAND-COVER AND LAND-USE. CASE STUDY : POST-COMMUNIST ROMANIA

Simona Niculescu¹, Philippe Fournier², Alexandru Badea³

1 : Laboratoire Géomer - UMR 6554 CNRS, Technopole Brest Iroise Place Nicolas Copernic, 29470 Plouzané
simona.niculescu@univ-brest.fr

2 : Service Central des Études et Enquêtes Statistiques (S.C.E.E.S.)
Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, France

3 : ROSA (Romanian Space Agency) : 21-25, Mendeleev st., Sect.1, 010362 Bucharest, Romania

Résumé

Les deux dernières décennies ont été marquées par une amélioration sensible de la production de statistiques agricoles en tenant compte du fait que la demande en information géographique est toujours croissante et de plus en plus diversifiée tant à l'échelle nationale qu'internationale. Cet article présente quelques aspects méthodologiques (les méthodes aréolaires conventionnelles) de production d'information géographique à travers les statistiques spatiales agricoles en utilisant les cartes topographiques à échelle fine et les images satellites à haute résolution spatiale comme base de sondage pour l'application des enquêtes agricoles en Roumanie post-communiste. L'article présente l'application du projet MARS en Roumanie en s'appuyant sur des exemples du Plateau de Falticeni, au nord-est du pays, et en focalisant sur l'adaptation de la méthodologie et de la collecte des données à une situation foncière et environnementale spécifique à un pays de l'Europe Centrale et de l'Est.

Mots clés : Information géographique, statistiques spatiales agricoles, images satellites à haute résolution spatiale, suivi et gestion du territoire.

Abstract

The last two decades substantial improvements in the production of agricultural statistics have occurred in a context where the demand for geographical information is constantly increasing and becoming more diversified at both the national and international levels. This paper presents some methodological aspects (conventional area-based methods) related to the production of geographical information via agricultural statistics using fine-scale topographical maps and high spatial resolution satellite images as sampling frames for the application of agricultural surveys in post-communist Romania. The paper describes the application of the MARS project to Romania based on examples from the Falticeni Plateau, North-East of the country, and focusing on the adaptation of the methodology and data collection to specific land and environment conditions in a Central and Eastern European country.

Keywords : Geographical information, spatial agricultural statistics, high spatial resolution satellite images, land monitoring and management.

1. Introduction

1.1. Agricultural statistics and remote sensing

Agricultural statistics-based systems are important tools for land and agricultural resources management. By determining the status of agricultural practices and production each year, agricultural surveys provide decision makers and land managers with viable and updated information concerning the agricultural environment. Contemporary agricultural statistics address issues such as estimating useful agricultural or nonagricultural areas, the distribution of cultivated areas at various scales (regional, national), forecasting potential crop yields and documenting year-to-year evolution and identifying general

trends.

Relying on agricultural inquiries based on area sampling¹, agricultural statistics are frequently used for the collection of data related to land-cover and land-use. Unlike mapping studies such as the CORINE Land-Cover project, area sampling (by point and segment) is a statistical method (EUROSTAT² 2000). The observation of various samples enables the establishment of primary crop regional area and yield estimates, considered as an acceptable generalization rendering the examination of the

1. Area bases are part of the hierarchy of geographical units (the subsections of the survey basis of a level can be subdivided in order to form units of a next level).

2. Statistical Office of the European Community.

entire territory unnecessary.

Recent developments in data collection tools (e.g., registries and administrative information sources, advanced sampling techniques and remote sensing) are at the forefront of developments in agricultural statistics. Remote sensing is a source of comprehensive and repetitive geographical information that can be used for agricultural statistics in association with conventional tools³ for the investigation of land-use, thanks in particular to new generation satellites with different resolution capabilities that are better adapted to the characteristics of European land plots. Remote sensing and agricultural statistics share a common field, the observation of land related phenomena. This technique, since the launching of the first Earth observation satellite (LANDSAT) in 1972, has contributed to the estimation of main crop areas and yield forecasting (among the better known applications). After 40 years of applying space borne remote sensing data to the field of agricultural statistics, an improvement can be observed in the classical methods of regional crop inventories, rapid area and potential yields forecasting based on zoning and the stratification obtained by remote sensing (area sampling, associated enquiries), leading to a better localization of the main land-use classes and to the systematic monitoring of crops, the state of the vegetation, the yield index and soil monitoring (MARS program). The classified satellite images (specially LANDSAT and SPOT and, in the USA the primary imagery source for deriving the 2009 CDL is the Indian Remote Sensing satellite IRS-P6, also known as Resourcesat-1 (Johnson and Mueller, 2009) are generally used as auxiliary variables, passing through regression (post-stratified and variance related) or calibration estimators based on confusion matrices. These registration and regression estimators combine accurate and objective observations related to the sample. Certain biophysical variables related to vegetation cover are accessible using remote sensing in different spectral domains. In particular, the leaf area index and chlorophyll concentration can be estimated using remote sensing data thus providing information relating to the potential crop yield. Remote sensing also reduces the number of sampling units required for land survey purposes while preserving, at the same time, the optimum level of accuracy for area estimates of the various themes found. Remote sensing is also used for the monitoring of crop evolution in a number of countries; for this type of application, repetitive data such as NOAA-AVHRR (National Oceanic and Atmospheric Administration - Advanced Very High Resolution Radiometer) data, images covering the surface of the globe twice a day, are equally useful as well as MODIS (Sun et al., 2008).

3. The data are provided by the general agriculture census, via specific inquiries of variable periodicity by a varied network of correspondents.

1.2. Study background

Several programs or international projects have been implemented for the introduction of remote sensing data into conventional methods to strengthen and improve agricultural statistical systems such as the American LA-CIE (Large Area Crop Inventory Experiment) program for crop forecasting and the AGRISTARS program (Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing), dealing with the improvement of agricultural statistics within the United States (Remote Sensing Applications in Agriculture at the USDA National Agricultural Statistics Service, 2010), AGRIT (Agrometeorology in Italy), which focuses on large crop yield forecasting, the European Union MARS and LUCAS projects with regard to the application of remote sensing to European agricultural statistics and the AGRIMA program in Morocco. These projects have been developed in the context of the increasing need for land-related information following the successive broadening and application of the new European common policy. In the framework of the implementation and monitoring of the agricultural policies of the European Union, the MARS (Monitoring Agriculture Remote Sensing) program was initiated at the outset. The expertise elaborated within this project refers to the use of remote sensing (satellite and airborne), geographic information systems (GIS), the global positioning system (GPS) and more general geomatics techniques. Recently in the USA, the National Agricultural Statistics Service (NASS) released a crop-specific land-cover classification product encompassing the entire conterminous United States (USA). Termed the Cropland Data Layer (CDL), the product depicts type and location for crops planted during the summer 2009 growing season. The remote sensing acreage estimation research program at NASS has since evolved paving the way for the current, more operational, CDL products. NASS has utilized AWiFS data for several years and finds it comparable to 30 meter Landsat-5 Thematic Mapper (TM) data for agricultural monitoring (Johnson, 2008). In order to contribute to policymaking, in 2000 Eurostat launched a second pilot project called LUCAS (Land-Use/Cover Area frame statistical Survey), which dealt with the application of area-based techniques and remote sensing to agricultural statistics. The identification of changes in land-use and landscape structure is an essential aspect mentioned in the project objectives. The environmental components, as well as the analysis of the interactions between agriculture, environment and rural areas, are given increased importance in community policies. The results of all these projects should be integrated into an advanced information system on agriculture, based on new information (remote sensing data at high and very high resolution) and more elaborate interpretation methods (agrometeorological models). The objective of this system is to test the integration and interactivity of the various conventional and new methods of remote sensing as an agriculture management tool (Gallego et Bams, 2008).

In line with these new European trends, Romania passed from the "pyramidal"⁴ system (exhaustive accounting type) applied to agricultural information before the fall of communism to the adoption of market economy rules during the post-communist period. This article presents the application of some of these methodologies to the territory of post-communist Romania and establishes correlations between local/regional results and the national and European levels. The objective of this paper is to show more specifically the implementation of an agricultural information basis, knowing that the information regarding the pilot projects is derived, most often, from a macro-regional even continental level, derived from a series of EUROSTAT type syntheses, where all the different levels are mixed up, which leads to the dilution of regional/local information. This article is an example of construction of the information on a regional scale.

2. Methods

Based on the application of principles of the MARS program in the Eastern countries, this contribution presents methodological elements concerning agricultural surveys based on closed segment stratified area sampling using small-scale topographical maps, as well as SPOT (10m x 10m) and LANDSAT satellite images (30 m x 30 m). The operational applications are provided by the European Union, which, since 1992, produces some of its agricultural statistics by means of area sampling (Gallego, 1999) in the framework of the MARS project.

This method has been used as of 1988 when 5 sites were chosen each measuring about 20,000 km² : France (center), Germany (Bayern), Greece (Macedonia), Italy (Emilia-Romagna) and Spain (Castilla-Leon). These sites were chosen in relation with their representative agriculture and the duration of the cloud coverage (to obtain satellite images without clouds). In 1991, the tests continued for the first sites (except Germany) also including other neighboring regions. In 1992, a large part of Spain and Greece was covered by this methodology. Other regions/countries were also experimented : the entire Belgium and three regions from South England. Two areas which did not belong to the EU were chosen : the entire Czech Republic and initially only a few regions from the South of Romania (according to the Rapport Regional Crop Inventories in Europe Assisted by Remote Sensing : 1988-1993.).

The contribution of this paper, which is mainly methodological, consists in the description of the different changes carried out in order to make the methodology

4. During the communist regime, the "pyramidal" information management system relied on the basic agricultural unit specific to a centrally planned economy (cooperative farming and/or state-owned farming businesses), which was actually an accounting system with obvious personal subjective marks.

better adapted to the territory and land management conditions of post-communist Romania. The produced geographical information refers to crop area statistics specific to the month of June of each year⁵ and to crop yield estimates and predictions of the main crops (wheat and corn).

2.1. Conventional methods

Survey techniques can contribute considerably to meeting the flexibility and quick response requirements necessary for land-use inventories and the estimation of main crop areas and yields. Random surveys consist of the random extraction of individuals from a total population. There are several statistical methods of selection that provide forecasts of what the statistical population is based on the knowledge of sampled individuals, provided that these individuals are selected randomly⁶ (random survey⁷). Systematic sampling can be assimilated with elementary or simple random sampling (Desabie, 1971). If the selection is not completely random, no representative results can be derived with a computable error margin (Batesse et al., 1988). These methods are often "in opposition" with the methods that use a "list" type basis. The random survey method used in this paper is based on the principle that the sample must be established in an objective manner, without the intervention of any "personal factors", so that each element of the studied group can have a chance to be chosen and in a way that its chances can be determined with certainty. Random selection can be applied to several primary or secondary units : departments, arrondissements, districts, communes, urban islands etc., on the condition that the systematic selection (here aligned systematically, cf. Figure 1) can be seen as a stratified survey. The notion of stratification consists in regrouping in a same class individuals presenting similarities, so as to minimize variability between two different layers. In order to perform an a priori stratification, knowledge of the population is essential to fit the individual into a single layer according to a given selection criterion⁸ (Cochran, 1977). This implies a very good knowledge of the land based on air photos, satellite images, maps or land surveys. Stratification often corresponds to the objective of reducing survey costs or optimizing its management. The main objective of stratification is to improve overall information accuracy for the whole set of elements involved.

For improving statistical accuracy, the stratified survey (cf. Figure 2) is recommended by many authors :

5. Compared to the classical agricultural statistics performed each Fall.

6. In the probabilistic sense of the term.

7. A survey is random or probabilistic if a whole statistical unit has a non-null and known probability of being selected in the sample (Ardilly, 2006).

8. In the United States, area stratification is carried out by taking into account the percentage of agricultural area in a given sector : from 0% for urban areas, forests, water. . . up to 75% for cropland in agricultural areas.

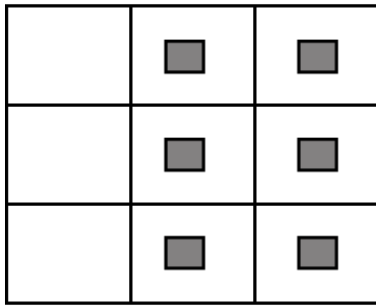


FIGURE 1 : Aligned systematic sample.

"stratification, even though imperfect, is preferable to the total lack of stratification" (Desabie, 1971).

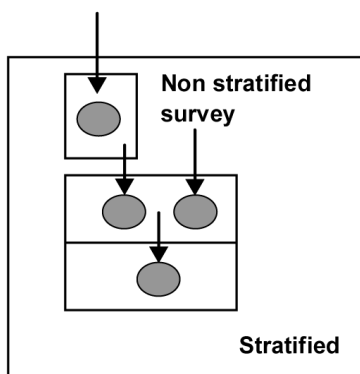


FIGURE 2 : Stratified survey principle (according to R. Clairin and P. Brion, 1997)

On the other hand, some authors say the contrary and feel that further stratification should not be applied with the a priori elimination of the land classes that were not previously estimated, thus risking the overestimation of agricultural areas that will be difficult to extrapolate. In the United States, physical limits (which can be layer limits) are used (Carfagna and Gallego, 2005). In France, the Ter-uti agricultural survey abandoned the notion of layers after three years; it was impossible to manage because of the borders between layers, departments, extrapolation coefficients, balance coefficients, etc. Since the 1960s, France has focused on a system of land-use surveys using point sampling (cf. Figure 3), while the United States has implemented a system of agricultural statistics surveys based on areas or "segments". In the case of land-use, point sampling has definite advantages when compared to area or segment sampling (Särndal et al., 1992); actually, the point is an essential component of the yield survey and the surveys relating to the new observation programs of the European territory (LUCAS⁹).

9. Land-Use/Cover Area frame statistical Survey. In 2001, the first Lucas pilot survey was conducted in thirteen member states of the European Union.

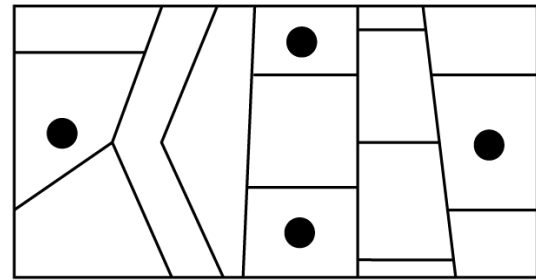


FIGURE 3 : Random area-based method using a point grid (Fournier, 1986).

The "segment" is not a portion of a straight line but "a portion of territory composed of contiguous parcels" (Fournier, 1986). Using a neologism borrowed from the Americans, we call "segment" an area of 500 ha (1 km wide in the east-west direction and 500 m high in the North-South direction)¹⁰. The segment method is an area-based method, which uses the subdivision of the territory (a "segmentation", hence the name of the method) into segments that are determined from natural limits (water courses, roads, paths, forests, railroads, etc.) or "artificially" built structures on the maps, often under the form of squares. Using random selection, statistical laws (probabilistic approaches) can be derived permitting the extrapolation of the survey results, the shifting of the sample (of the closed segment) to the study area and, also, the calculation of the accuracy (for further details, see the third section).

The closed segment¹¹ method consists of the identification of all the land-use types within a segment in order to produce area estimates. This method refers primarily to the physical aspects of the segment composition. Since the observation unit is the segment, estimates are thus produced depending on the selection mode of the segment sample. The individual sample membership probability of the land-cover and land-use classes is thus proportional to the area that they actually occupy.

Apart from the survey, a sampling frame² (or survey support) is required to be able to go from the sample to the study area. The priority in the case of a sampling frame is its updating, given the fact that a "sampling frame becomes obsolete" (Brion, 1995). Moreover, a sampling frame must be adjusted to the study field and area. It is recommended to use a "composite sampling frame" (Brion, 1995): fine-scale topographical maps and satellite images. Thus, satellite images can constitute the survey sampling frame. Generally, these are used

10. According to the basic methodology, a segment must comprise 5 to 15 plots/parcels.

11. Apart from this method, also worth mentioning are: the open segment method, which consists in the analysis of all the agricultural exploitations found in one of the segments selected in the sample and the weighted segment method? which enables the designing of a questionnaire for each agricultural exploitation that has lands on the segment.

as a sampling frame for demographic surveys in urban environments, but also for agricultural surveys as comprehensive auxiliary¹² information or as support for the stratification of the territory into homogeneous areas¹³. Present studies in statistics show the importance of both the estimation of the variance and use of auxiliary variables when several samples are available. The information on the variance of an estimator enables the evaluation of its performance and the determination of its accuracy. The use in an "optimal way" of auxiliary information should permit the improvement of the quality of an estimator (Deville and Särndal, 1992 ; Särndal, 2007). The efficiency of remote sensing-based stratification varies according to the refinement of the photo-interpretation, but especially with regard to the distribution of the studied land-use. Stratification efficiency is often limited by "the multi-objective nature of the stratification and the associated sample" (Gallego, 2005). Otherwise, a more detailed mapping support can be used. Thus, mapping must be restricted to the visualization of the area distribution and extension of the various land-use themes, whereas the statistics enable a global and theme-based analytical approach, which mapping does not allow.

A last aspect concerning sample surveys is survey accuracy. The estimate obtained from the sample must be as close as possible to the unknown value to be assessed. For this reason, to improve results a probabilistic sampling method¹⁴ should be used since "only probabilistic surveys enable the estimation, at the end of the acquisition process, of the sampling error and, thus, the possibility of acting on the sampling plan to reduce it" (Barbary, and Dureau, 1991).

3. Examples of area (spatial) sampling

Transition, in the former Eastern countries, can be defined as the process of establishing new operating rules for the economy and society. For statisticians, it refers to documenting the new environment pertaining to the economic and social reality based on adequate indicators (Suesser, 1997). In these countries the overall situation in agriculture, as well as in agricultural statistics, is more traditional, and at different stages en route towards the EU model. The expansion of the EU in 2004 and 2007 and the enhanced cooperation with future member countries and other neighboring countries

12. Information concerning statistical units, other than their location, which can be used in the survey ; these items of information are called auxiliary variables that can be used to improve the selection technique or for the calculation of a more efficient estimate (Chikara et al., 1986).

13. Homogeneous areas in relation to a variable supposed to be (or known to be, following a comparative study) highly correlated with the one(s) to be estimated.

14. Methods of sample selection based on randomness and random selection ; the method is said probabilistic because the representativeness of the sample is assured a priori by the statistical laws of probability.

have visibly increased the impact of decisions on the organization and content of agricultural statistics in the EU. The greater variety in crops and methods demands a review of existing statistics, particularly in the EU context.

3.1. Area sample surveys

In Romania, the new statistical information system was built under time constraints, with a rather clear vision of the main objectives and many new instruments to install ex nihilo. During the 1992-2004 period, two European projects pertaining to European statistics were carried out in Romania (MARS and LUCAS). The two programs relied on main crop area sample surveys.

The sample survey relies on the field study of a certain number of rectangular shape areas ($500\text{ m} \times 1\text{ km} = 500\text{ ha}$ ¹⁵), without physical borders, geographically distributed over the territory taken into consideration (in this case, the territory of Romania). The objective of this application is to provide estimates at the national level of the main crop areas with an adequate level of accuracy, and to subsequently provide, at cropping time, yield forecasts for crops of major interest. With regard to the initial objective, it was difficult to obtain satisfactory statistical results at the level of each of the counties in the country¹⁶. The survey actually provides global statistics concerning land-use and all the types of land-uses have a probability of belonging to the sample, proportionally to its areal extension. This area (the closed segment) without physical limits in the field is first marked on the 1/25 000 topographical maps covering the whole country, according to the selection principle presented below. Based on the experience of other countries (USA - EU) and the experience gained in the Eastern countries since 1992, a sampling fraction of 1% of the territory was established. Having retained the rectangular shape segment with an area of 500ha, theoretically Romania is covered by 476 782 segments, from which we extract 1 from 100, or approximately 4,768 segments at the country level (cf. Figure 4, a North-East portion of the country, and the Falticeni plateau (Niculescu, 2002)).

The 1/25 000 topographical map of Romania includes the superposition of a kilometric grid ($1\text{ km} \times 1\text{ km}$) in the East-West and North-South direction, representing a partitioning of the territory of the country into square grids of 100 km^2 . The grid is 2 km by 2 km on the 1/100,000 maps. The survey consists in the selection of two 0.5 km^2 segments within each 100 km^2 grid to ensure the desired sampling fraction. There are several selection methods within the grids, all of which ensure, to a greater or smaller extent, the geographic dispersion of the sample (Kish, 1994). Two independent

15. The segment sizes were not modified in relation to the American model, even though the segment sometimes included up to 43 small plots. The ultimate goal was to ensure a significant variance within the segment.

16. Romania is divided into 42 counties or « judete » (in Romanian).

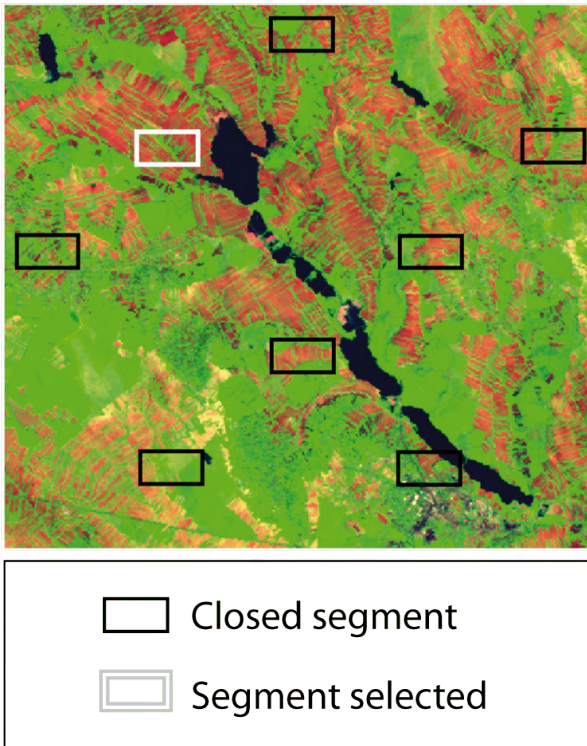


FIGURE 4 : Closed segments on the northern side of the Falticeni plateau (North-East Romania, LANDSAT image (30 m × 30 m)).

selections, both systematic, were made following the East-West and North-South directions (Niculescu, 2002).

Systematic selection consists in selecting individuals whose ranks are in arithmetical or equidistant progression in relation to the other in the case of area-based surveys. The first selection is called "repetition" 1 and all the segments of this "repetition (Independent probabilistic sample)"¹ are equidistant at 10 km according to two privileged directions, and the second selection, which is not related to the first one, is called "repetition" 2 (cf. Figure 5).

The mapping projection is not identical over the entire territory of Romania, the separation being aligned East-West. To avoid any overlapping or shortcoming on the North-South band, which is of interest to us, either the repetition 1 segment or the repetition 2 segment were retained to maintain the geographical distribution of the sample. As the choice is made on the map, it is purely random and does not favor a specific land-cover or land-use. The final selection for Romania produced 53 sampling units (from a total of 4 766) insuring a sampling fraction of 1/99.96 and an average extrapolation coefficient of 100.04 per segment. Hence, the selection carried out meets the requirements of randomness and geographical distribution, respecting the sampling fraction established a priori. The segments were numbered successi-

vely, from 0001 to 4766, at the national level, line by line, from west to east and from North to South. The numbering does not take into account the "initial" stratification performed at the country level, since this would imply the reintroduction of the results obtained on "layer 1" into the results obtained from the other layers.

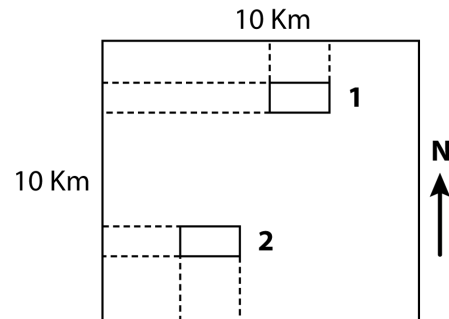


FIGURE 5 : Diagram showing the selection of two segments within a grid square.

The numbering from 0001 to 4766 implies the fact that the focus is on a land-use survey rather than on a simple agricultural survey since all the "segments" theoretically participate in the result of the survey. Along the same lines, the segment numbers of repetition 1 (or repetition 2) are successively arranged, unit by unit. Actually, the segments have received a double numbering, following the same principle (West-East/North-South), but at the level of each county, implying that the overall segment was allocated to a single county : this can be explained by the fact that the borders between counties are administrative limits and not natural ones¹⁷. In France, sampling relies on 8 possible repetitions within a 12 km × 12 km = 144 km² grid, whereas the Romanian model is based on 10 possible repetitions within a 100 km² (10 km × 10 km = 100 km²) grid.

When partitioning the universe¹⁸ into more or less homogeneous areas based on conventional topographical maps or satellite images, the term zoning (In remote sensing, areas of equal appearance) should be used instead of stratification. Using topographical maps and satellite images, a zoning introducing the notion of "layer 1" was performed, including the mountain area (Carpathians) and the Danube Delta. This "layer 1" is covered by 1,120 segments at the level of Romania as a whole. The forests growing in the plains and large cities were not isolated from the agricultural layer.

Overall, over the territory of the country, an area selection

17. Experience shows that the management of samples overlapping two administrative units or for two different layers is practically impossible (with reference to agricultural statistics in France).

18. The groups studied by the Descriptive Statistics bear the general name of statistical universe or population.

of geographically distributed 500 ha units was carried out following the two independent selections made at every ten kilometers in the East-West and North-South directions (cf. Figure 5). The allocation of the "segments" of the national sample is by county and introduces "layer 1", characterized by little agricultural activity.

The nomenclature used in this survey is a physical land-use nomenclature, with some functional notions for the non-agricultural dimension. This nomenclature covers all the land-use types (perhaps it could be improved at the level of the "natural environment", forest being susceptible of regrouping all the areas of spontaneous woody vegetation and natural pastures regrouping all the heaths and wild lands). There are 75 nomenclature themes, 41 of which refer to arable lands from 61 agricultural themes. Another important aspect of surveys is the issue of surveys made by pollsters. This is the key aspect of the survey. The random error due to the fact of pursuing the survey in the first place can be easily calculated using the appropriate statistical formulas for the type of survey used. The observation error is more difficult to compute, as there are several types of possible errors : location error, error in the tracing of the parcels (omission, for example), error concerning crop nature. The total error is the result of the two types of errors. Whatever the improvements brought to the survey (multiplication of the segment number, stratification or finer zoning), if the observation error is important, the total error will also remain important.

3.2. Pollster field observations

At the time, the air photo coverage of Romania was not available or too old. The cadastre has not been updated since the land redistribution (1991) ; thus, it cannot be used. The only support available was the enlargement to 1/10 000 of the 1/25 000 topographical map and/or high resolution satellite images (here, SPOT P, 10 m × 10 m). The maps were not very recent but had a sufficient number of landmarks and present an adequate sketch of the road system, so that a field line can be drawn without too many errors. Since the 1/25 000 map was the support for sample selection, there is no error in transferring the segment limits with the sides well oriented East-West or North-South.

In Romania, the topographical map and the SPOT (10m × 10m) and LANDSAT satellite images (30 m × 30 m) constitute the basic survey documents. The samples presented are from the Falticeni Plateau located in North-East Romania (Niculescu, 2002). The 1/25,000 map provides information on how to approach the segment and what itinerary can be predicted (cf. Figure 6). As mentioned before, the map enlargement to 1/10,000 and the SPOT (10 m × 10 m) image are the survey documents used (cf. Figure 7). The use of satellite images is definitely beneficial. The information provided relating to the parcel plan structure is considerable despite potential confusion related to the identification of geographical

objects on the ground. By comparison, a survey file elaborated from points or segments gives only the values at the level of those entities and provides no information with regard to their interrelations.

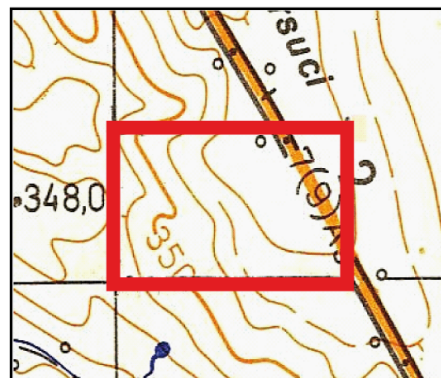


FIGURE 6 : The topographical map permits the defining of an access path (1/25,000).

On this site survey at 1/10,000 (cf. Figure 8), 1 mm represents 10 m and it is difficult to expect a relative accuracy in the field of 3 to 5 m. To draw a parcel with its borders, it is essential that the parcel be at least 12 to 15 m wide. The drawing error can be important : it is recommended for the pollster to produce a working document of the field before actually elaborating the final version for the parcel plan on the overlay. The joined micro parcels are regrouped under a single number with the code "-2"¹⁹. The parcels are numbered and coded by crop type by the pollster himself. For the parcels coded "-2", the pollster records each crop found and calculates the number of steps he takes across every field and then transcribes these data in a specific record.

3.3. Results of area estimates

Results in this section refer to statistical area estimates and a few indicators (variance, standard deviation and coefficient of variation), which are necessary for result interpretation.

$$Sh = S \frac{1}{m} \sum phi, \quad (1)$$

Where :

- Sh = area in category "h" territory to be estimated ;
- S = area of Romania²⁰ = 239,391km² ;

19. This is a conventional code.

20. According to official statistics, it is distributed into : 147,931 km² of agricultural areas of which : 93,415 km² are arable land, 33,626 km² natural prairies and pastures, 14,893 km² meadows for harvesting, 3,039 km² vineyards, 2,958 km² orchards ; 66,810 km² forests, 8,926 km² water, 14,724 km² other types of lands (buildings, transportation infrastructures, etc.). Agriculture and forestry represent more than 62% of the national territory (90% together with forests). Agriculture and forestry represent over 20% of the national gross added value.



FIGURE 7 : On the SPOT P image (10 m × 10 m) the same "segment" is located in relation to its environment, Falticeni Plateau.

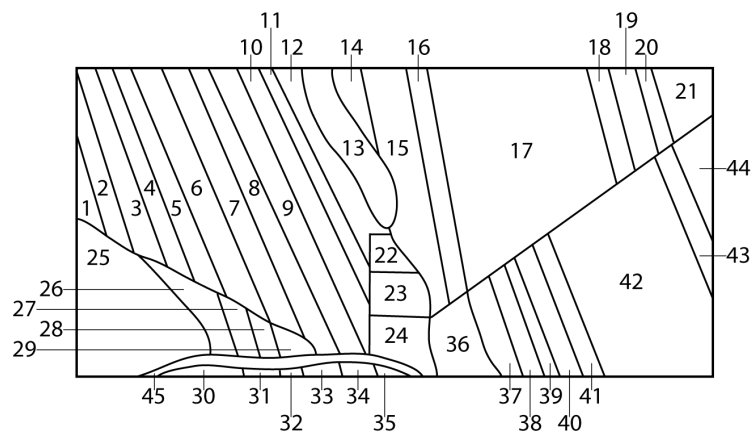


FIGURE 8 : Site survey of the same segment from the Falticeni Plateau.

Crop	Mean (ha)	Mean (%)
Wheat	18,904,340	4.624
Barley	720,718	0.176
Variety of barley	3,504,781	0.055
Oat	13,592,216	3.325
Rye	2,443,665	0.592

TABLE 1 : Example of statistics obtained for Suceava country (North-East of the country) in June (1998).

- m = total number of selected segments ;
- i = number of segments ;
- ϕ_i = percentage of category " h " territory within the segment " i ". ϕ_i is computed by summing up the numbering of the segments : the areas of parcels that do not have a "-2" code and by allotting to each field regrouped into the "-2" class an area proportional to the number of steps made by the pollster.

The variance of this estimator is :

$$V(sh) = S^2 \times \left(\frac{M-m}{M} \times \frac{1}{m-1} \right) \times \sum (\phi_i - \bar{\phi}_h)^2. (2)$$

M is the theoretical number of segments covering Romania ($\frac{M-m}{M} = 0.99$ is very close to 1 and can be neglected). $\bar{\phi}_h$ is the average percentage in category h territory calculated for all the segments. Standard deviation is :

$$\sigma_h = \sqrt{V(Sh)}. (3)$$

The variation coefficient CV_h is :

$$CV_h = \frac{\sigma_h}{Sh}. (4)$$

It is a dimensionless variable expressed in %.
The survey was sorted by county (cf. Table 1) and by "-2" coded parcels, not proportionally to the number of steps taken, but according to the relative frequency of each of the crops present in these parcels (a crop is counted each time it appears in this parcel).

In proceeding this way, it can be argued whether the agricultural is not favored in relation to the non-agricultural land-use. How are paths, road or railroad networks, drainage or irrigation canals treated? The segment survey refers to gross agricultural areas, whereas the point survey deals with net agricultural areas (all land-uses 3 m wide or with an area of more than 10 m² has the probability of being selected, and area estimates can be derived for roads, isolated trees, trails, etc.).

3.4. Yield sample surveys

A grid of six equidistant points located at a distance of 350 m from one another in the field is superimposed on the segment plan (cf. Figure 9). Starting from the point going to the farmer who is responsible for cultivation of the field, an interview is conducted with regard to certain parcel characteristics and main crop yields at the level of the overall cultivated area.

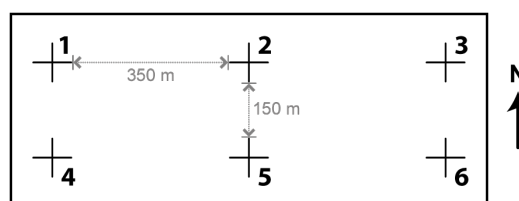


FIGURE 9 : Yield sample survey.

Only one out of three segments for this yield survey was retained, namely segments that can be divided by 3 in the national number (whatever the repetition). This selection guarantees the dispersion of observations. Only the segments of layers 2 and 3 (except for layer 1) are part of the sample. The plain segments with 100% forest selected in the yield survey are visited by the pollster to verify whether there are crops on the segment. The interview with the farmer will only take place after harvesting and, then the pollster will go back in the field : in the case of large parcels where the yield selection point can be attributed without ambiguity, the pollster can obtain information concerning the number and the address of the farmer starting in June. Each of points is equiprobable and all have the same representativeness. The information obtained at the parcel level and correlated with the area unit is self-weighted : the estimate of an average will be the average of the identified criteria, correlated with the area unit²¹. A parcel must be counted each time it has been selected (as many times as there are points). The parcels have a probability of being selected proportional to their area (unequal probability survey). First the exploitation area must be defined : by taking into consideration only the area of arable lands ; or the area of arable lands, plus orchards and vineyards, thus the agricultural area used (SAU) ; or the useful agricultural area (SAU, plus heaths and wild lands). The data collected at the level of each agricultural exploitation is weighted by the reverse of the exploitation area. Selections with unequal probabilities require that there is a good correlation between the studied variables and the selection variable. Here, the selection variable is the exploitation area. Production is related to the area and not to the crop distribution.

3.5. Results of yield estimates

Given the fact that the variables to be studied must be correlated with the selection variable, only the main crops are of interest (in France, the survey concerning tillable lands is limited to only 6 to 7 crops²²).

21. Example : average dose of agricultural fertilizers = average of the dosages used for 1 ha.

22. In France, approximately 300 agricultural exploitations per county are retained. For the smaller counties in Ile de France, the number decreases to approximately 150 exploitations (but tillable lands represent almost 95% of the agricultural area used).

The crop by crop yield is obtained through the ratio between production and area (the two are estimated from the exploitation records) (Fournier, 1986) Eventually, an accuracy ratio of this estimator²³ can be calculated. In Romania, the survey is conducted by taking into consideration the exploitations generated by the points selected over the entire area of the territory. If a point (with a null theoretical area) is selected on a territory with the area "S", the probability of this point including exploitation "I", which means that it falls within the area occupied by this exploitation, is definitely proportional to the exploitation area s_i ; this is precisely s_i/S . The selection is a two-degree selection and systematic unequal probability.

3.6. Discussion

The advantage of random methods resides in their ability to estimate result accuracy. Theoretically, they require prior knowledge of the area under survey, which can be limited to the sampling design support acquisition. Airborne or satellite remote sensing provides excellent land survey support, namely high spatial resolution images, but also wide field-of-view images. Due to their comprehensiveness, these data are quite useful as sampling design support. The major constraint resides in the quasi obligation to investigate the selected individual - object -, so as to minimize the risk of the sampling losing its randomness (Desabie 1971).

The implementation of the MARS project in Romania between 1992 and 1998 shows that the USA method is appropriate for the specific territory and land management conditions of post-communist Romania. First among these adaptations is the number of plots included in a closed segment. In the basic methodology, a closed segment must include between 5 and 15 plots, while in Romania the segment can comprise up to 45 plots, taking into consideration the fact that the land was redistributed to its former owners. The political changes that occurred in 1989 in Eastern Europe led to significant changes in the centrally planned economy, which had a considerable impact on the evolution of agriculture. The passing of the law regulating land redistribution in Romania resulted in a higher number of landowners, which fundamentally altered the structural agricultural unit itself, from a centrally planned economic system model to a dramatically changed land base structure.

The application of the closed segment method in Romania also took into consideration the characteristics and topographical map coverage of Romania; considering that the map projection is not identical throughout the entire country; a systematic sampling from North to South and from East to West was proposed. Ten possible repetitions within a 100 km² grid were carried out with reference to the administrative limits of the 42 counties in

23. The average yield can be estimated by calculating the yield arithmetic mean (each yield is taken into account as many times as the exploitation was selected). When dealing with large numbers, the result is slightly different from the one calculated with the help of the rigorous formula.

Romania.

A zoning introducing the notion of layer 1 (lightly agricultural) was proposed taking into consideration the mountain area (Carpathians) and the delta area (Danube Delta). Hence, a land-cover and land-use nomenclature was produced in relation to the geographical stratification of Romania. Finally, the size of the closed segment was chosen depending on the plot fragmentation in the "plain" layer, in order to achieve significant segment variance. For insuring the harmonization of the European agricultural system, after 1998, at the suggestion of the EU, Romania implemented area sampling methods based on a point grid in the framework of the LUCAS project. Theoretically, observations are performed using a dimensionless point. Thus, proportions are normally calculated rather than areas. In parallel, agricultural statistics are also produced based on the agricultural areas declared by farmers applying for subsidies from the EU and the Romanian Government. Subsidies are granted in accordance with European regulations (IACS = Integrated Administration Control System. LPIS = Land Parcel Identification System.). The system is based on the geographical plot location principle. Farmers listed in the official records have to declare the area of their different crops to receive European and national subsidies. There is a control system to verify these declarations that relies precisely on geographical plot location using satellite images and GPS ground tracking. The Agricultural Statistical Network Department is subordinated to the Ministry of Agriculture, Forests and Rural Development (MAPDR) and cooperates with the National Statistics Institute (INS).

4. Conclusions

This operation permits the testing in real size of a system combining satellite images with data collected by the statistical services. The feasibility of the system provides for the reduction of the number of satellite images to be acquired and processed to a minimum. This imposes that we work on a reduced number of sites at the limit of what statistical theory considers a random sample, which is amplified by geographical continuity or discontinuity phenomena that are difficult to handle, especially the concentration of certain crops on a reduced number of sites. The main difficulty in yield surveys consists in the fact that the information must be collected throughout the entire year as crops mature and certain data become available. To preserve the timely nature of these estimates, publication of the results before having access to the total number of questionnaires must be contemplated. Waiting for the last questionnaire would lead to delays similar to those of large national surveys.

The use of satellite images as survey documents does not permit the estimation of areas of certain land-use categories that have a linear or punctual shape characterized by dimensions inferior to the satellite resolution. The

use of the latest-generation metrical and sub-metrical satellite images (IKONOS, QUICKBIRD, etc.) in agricultural surveying promotes point area sampling. The diversity of error sources excludes the implementation of a method that systematically corrects these errors and leads to the search for a global correction method²⁴. Recent studies (Everaers, 2010) conclude that globalization and issues such as climate change require a different approach to statistics, given the important role of agriculture in the global economy, global economy sustainability and contemporary society more generally, and this clearly includes agricultural statistics. However, the interrelatedness of agriculture with, for example, land-use and rural development, and also with environmental sustainability and overall wellbeing, is considered as being not yet fully reflected in available statistical information.

Références

- Ardilly P., 2006. Les techniques de sondage, Paris, France, Technip, 393 p.
- Bailey J. and Boryan C., 2010. Rapport Remote Sensing Applications in Agriculture at the USDA National Agricultural Statistics Service, USA, 14p.
- Barbary O. and Dureau F., 1991, L'enquête par sondage sur image satellite : une solution pour améliorer l'observation des populations citadines. STATECO, 67, pp. 44-58.
- Battese G.E., Harter R.M. and Fuller W.A., 1998. An error-components model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association, 83, pp. 62-98.
- Brion P., 1995. Base de sondage : entre rigueur et bricolage. Clins d'oeil de démographes à l'Afrique et à Michel François, n° 2, Paris, CEPED, pp. 117-124.
- Carfagna E. and Gallego F.J., 2005. Using Remote Sensing for Agricultural Statistics. International Statistical Review, 73, pp. 389-404.
- Chikara R.S., Houston A.G. and Lundgren J.C., 1986. Crop acreage estimation using a LANDSAT-based estimator as an auxiliary variable. IEEE Transactions on Geoscience and Remote Sensing GE-24(1) :165-198.
- Clairin R. and Brion P., 1997. Manuel de sondages. Application aux pays en développement, n°3, Paris, France, CEPED, 108p.
- Cochran W., 1977. Sampling Techniques. John Wiley and Sons, Third edition, New-York, USA, 24p.
- Desabie J., 1971. Théorie et pratique des sondages, Paris, France, Dunod, 481p.
- Deville, J.C. and Särndal, C.E.. 1992. Calibration estimators in survey sampling. Journal of American Statistical Association 87(418) :376-382.
- Everaers P., 2010. The present state of agricultural statistics in developed countries : situation and challenges. Agricultural Survey Methods, John Wiley & Sons, New-York, USA, 24p.
- EUROSTAT, 2003. L'enquête LUCAS. Les statisticiens européens assurent le suivi du territoire. Technical report of the European Union, 24p.
- Fournier P., 1984. La télédétection au service de la statistique agricole. Travaux de l'Institut de Géographie, pp. 49-63.
- Fournier P., 1986. Enquêtes sur l'utilisation du territoire effectuée en 1985 par la méthode des segments, Paris, France, SCEE, série S, 13, 95p.
- Gallego F.J. 1999. Télédétection et estimation des superficies. Ed. Sc. G. Brossier et A.-M. Dussaix, Enquêtes et sondages. Méthodes, modèles applications, nouvelles approches. Paris, France, Dunod, pp. 60-68.
- Gallego F.G., 2005. Stratified sampling of satellite images with a systematic grid of point. ISPRS Journal of Photogrammetry and Remote Sensing 59(6) :369-376.
- Gallego, J. and Bamps, C., 2008. Using CORINE land-cover and the point survey LUCAS for area estimation. International Journal of Applied Earth Observation and Geoinformation 10(4) :467-475.
- Johnson D.M. and Mueller R., 2010. The 2009 Cropland Data Layer. Photogrammetric Engineering & Remote Sensing 76(11) :1200-1205.
- Johnson, D.M., 2008. A comparison of coincident Landsat-5 TM and Resourcesat-1 AWiFS imagery for classifying croplands. Photogrammetric Engineering & Remote Sensing 74(11) :1413-1423.
- Méthodes de sondage pour les enquêtes statistiques agricoles, 1994. Collection FAO : Développement statistique. Organisation des Nations Unies pour l'alimentation et l'agriculture, Roma, Italy, FAO, 60p.
- Meyer Roux J., 1990. Le Projet pilote de télédétection appliquée aux statistiques agricoles en Europe. La télédétection en agriculture, Série A / n°4, pp.57-63.
- Niculescu S., 2005. Transformation des territoires ruraux (Editorial). Revue Internationale de Géomatique 15, 10p.
- Niculescu S., 2002. Approche géographique de la dynamique des paysages du Plateau de Falticeni (Roumanie) par télédétection. PhD thesis, Paris IV, France, 396p.
- Särndal C.E., Swenson B. and Wretman J., 1992. Model Assisted Survey Sampling, New York, USA, Springer Verlag, 320p.
- Särndal C.E., 2007. The calibration approach in survey theory and practice. Survey Methodology 33 :99-119.
- Suesser J.-R., 1997. Les systèmes d'information dans les pays de l'Est. Transition statistique ou statistique de la transition ? Statéco, 85, Paris, France, Insee, pp.5-12.
- Sun W., Liang S., Xu G., Fang H. and Dickinson R., 2008. Mapping plant functional types from MODIS data using multisource evidential reasoning. Remote Sensing of Environment 112 :1101-1024.
- Taylor J.C., Sannier C., Delincé J., Gallego F. J., 1993. Ispra Rapport Regional Crop Inventories in Europe Assisted by Remote Sensing : 1988-1993, 1997, 64p.

24. This method of global correction relies on the allocation of sites to virtual layers defined by a given percentage of tillable lands. The notion of virtual layer, with unknown geographical limits, but for which the area can be estimated, enables the formalization and development of the initial notion of selection.