Computers and Electronics in Agriculture
20 (1998) 45–77

Computers
and electronics
in agriculture

Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas

T.A. Tsiligirides *

Department of General Science, Informatics Laboratory, Agricultural University of Athens, 75 Iera Odos, 118 55 Athens, Hellas, Greece

Received 12 July 1997; received in revised form 1 December 1997; accepted 30 December 1997

Abstract

The 10-year Monitoring Agriculture with Remote Sensing (MARS) project, launched by EU in 1989, initiated several studies with the aim of providing technical support to the European Agricultural Guidance and Guarantee Fund (EAGGF) and helping the Member States' Administrations to improve and industrialise their methods using remote sensing technology. One of the main targets of the project is to provide the framework for implementation of this technology assessing the use of high resolution satellite images as part of the measures to be taken by the national governments to improve the ground survey estimates based on an area frame rather than a holding-based approach. The area frame methodology is primary used for crop area, and yield or production estimates. The statistical units (the segments) of an area frame are directly bound to a stratified geographical region, the limits of which are known in advance. Thus, the elements of the population (the frame) are also known. Since the population is stratified no segment may be shared by two or more strata. Samples are obtained by dividing a region into blocks of equal (square) shape and repeating a pattern of elements across the region. In this context, two-dimensional systematic aligned sampling methodology, with a distance threshold in a stratified area frame on a square grid, is developed. The main purpose of this paper is to review the area frame of square segments methodology when it is applied in combination with satellite imagery (also known as supervised classification), and to provide the main features that appear when it is used in agricultural statistical surveys. Further, it is to report results of area estimates

* E-mail: tsili@auadec.aua.ariandne-t.gr

PII S0168-1699(98)00011-8

obtained from the implementation of this fast acquisition statistical data methodology, using Landsat-tm images in certain cultivated Hellenic areas, particularly productive in soft and durum wheat, maize, cereals, sugar beets, cotton, tobacco, olives trees and vines. Although the results show that there is some improvement in using the supervised classification methodology, a revised stratification methodology is proposed and a new sample is extracted for the Hellenic regions of Macedonia and Thrace, using no satellite data. The new classification is simpler, easier and less costly to implement than the one that is in current use. The developed regression model provided more accurate and viable acreage estimates than a previously applied model, and it may be extended to all cultivated Hellenic regions. Finally, the acreage, yield and production estimates obtained from the supervised classification methodology are compared with those obtained from the rapid estimates methodology which was also developed in the framework of the MARS project and is reviewed here. This comparison shows a noticeable agreement in the results obtained.

Keywords: Remote sensing; Area frame sampling; Agricultural statistics; Agricultural surveys; Crop estimates

1. Introduction

The reform of the Common Agricultural Policy (CAP) was devised as an attempt to reduce the over-production of certain crops within the European Union (EU), with subsidy payments for these crops based on area rather than production (de Winne, 1994). This change in the system requires individual farmers to submit area-based subsidy applications directly to the competent authorities, giving a breakdown of their farm by field, with area land use and geographical location clearly identifiable.

Traditionally, surface-based subsidies have depended on ground inspections. These are time-consuming, therefore costly. They also raise problems on quality control. On regional and EU scales large amounts of data have to be processed, in some cases by reference to data from previous years. In the case of annual crops the need to process the data before harvest imposes considerable time constraints. Under the reformed arrangements modification of techniques is therefore required, in the interests of time and cost savings, and to ensure that the new regulations will be implemented.

Satellite remote sensing (RS) provides synoptic, objective and homogeneous data which can be geographically and temporally registered, and therefore, could be an efficient tool for providing standard, high quality information on agriculture, evenly over the whole of Europe. The Monitoring Agriculture with Remote Sensing (MARS) project of EU was established in order to define and demonstrate how RS could be used operationally to supplement, interpret, and standardise agricultural statistical data provided by conventional techniques (Meyer-Roux and Vossen, 1993). The 10-year (1989–1998) project was initiated by the Directorate General for Agriculture (DG-VI), in co-operation with the Statistical Office of the European

Communities (SOEC). The Agricultural Information Systems (AIS) Unit of the Institute for Remote Sensing Applications (IRSA), a research organisation of the Joint Research Center (JRC) of EU in Italy, is responsible for implementing the programme, in close co-operation with national laboratories and organisations.

Initially, the MARS project concentrated on the production of statistical estimates on a regional scale. Then, using this experience, computer assisted photo-interpretation and geographical information systems (GIS) were introduced and the work was directed towards analysis at a field level, thus meeting the requirements of an area-based agricultural control scheme. In this context, the implementation of RS and GIS optimise the efficiency of ground controls, by guiding field controllers to non-conforming applications only so that reducing the time and therefore the cost of on-the-spot checks as only certain fields have to be measured. During this phase of the project (MARS-I; 1989–1993) various activities were conceived, developed, and implemented on the basis of inputs from approximately 100 institutions in 17 European countries. These institutions provided the required data, models, algorithms and software after having previously validated them for use at the EU scale on the basis of site, or country specific information. As an example, in 1993, 240 satellite images were processed to check 35000 applications in 44 sites over 11 Member States of the EU. The methods developed were improved and expanded further during the following years, bringing the cost-effectiveness, time efficiency and objectivity of RS and GIS to the support of the CAP reform and its associated area-based subsidy schemes. Main achievements from the implementation of these technologies were the harmonisation of the control procedures between Member States, and the assimilation of the information acquired into the Integrated Administration and Control System (IACS) that has been running from 1996 (activity G of the MARS-II project). In addition the information gathered into the IACS includes database systems for the management of applications and GIS for plot registration.

In order to meet these objectives the MARS-I project was organised in seven actions. The priorities set were the inventories of acreage, the inventories of production and the production forecasts. The introduced RS methods were to be tested on fairly large areas, such as states or provinces and to be developed to a stage where they could be put into operational use. This entailed the use of satellite data for which there would be guaranteed continuity. Pixel size (Landsat-tm 30 m, Spot XS 20 m, Spot Pan 10 m), set the limits of accuracy of field measurements from satellite data, technical tolerances were determined by the Commission according to various studies (Terres and Arvain, 1993). The crops targeted were those with the biggest market, excluding the crops consumed on the farm, such as fodder crops. Representativeness was sought not only at Community level but also at regional and national level.

However, in the second phase (MARS-II; 1994–1998) the project was reorganised and the principal objectives were classified into three major fields (subprojects). The first sub-project is MARS-STAT, which aims at improvement of the agricultural statistics, using RS methodologies and techniques. Its main objectives include quantitative estimation of the acreage occupied by the various crops in a given region or country; monitoring of vegetation and crop state; timely crop yield forecasting of mean crop yields for a given region, and finally rapid and timely estimation of the total production of the most economically important crops. The second sub-project is MARS-CAP, which aims at the support of CAP reform. It involves activities related to crop identification by satellite RS activities with a broader scope, such as the assessment of the precision and reliability of plot location limits, as well as the conceptual development of the IACS. The third sub-project is MARS-Extensions, with the objective of transferring methods and their adaptation to non-EU countries, with, in order of priority, the PHARE, the TACIS, and the MAGHREB countries. The MARS project undertakes in this context activities in the framework of the MERA (MARS and Environmental Related Applications) project, within the PHARE Regional Environment Programme. Within the same project, the extension of MARS activities to Latvia, Estonia, Lithuania, Albania and Slovenia was started in 1995. These activities will then be extended to other Baltic States and to Russia (Pertigao and Burrill, 1994).

The following concentrates on the development and implementation of two operational activities of the MARS-STAT project.

Activity A (Regional inventories): this consists mainly of the assessment of crop acreage with high resolution RS imagery (mainly Spot and Landsat-tm images).

Activity B (European rapid estimates of acreage and potential yields): this provides rapid estimates on the European scale of actual planted areas of the main annual crops, as compared to the previous season.

As the above descriptions suggest, both activities provide a common ground for comparison to the different methodologies used in order to obtain estimates on the acreage, yield and production of certain crops. Their nature determines the data-gathering method, which consists of surveys in which the sampling unit is an area, rather than an individual or a farm as is usually the case, hence the name area frame sampling given to these surveys. In order to facilitate the understanding of how the area frame sampling methodology is linked with RS a brief description of the main phases of the process will be given in the most common case, that is, when ground data are gathered as fields, or clusters of neighbouring fields, called segments.

There are three main objectives of this research. The first is to provide an overview of area frame sampling methodology. Emphasis is given to square sampling units or segments, as this method is much cheaper to implement and is more suitable for adoption in developing countries. This methodology has been applied in both A and B activities of the MARS-STAT project. The second objective is to show how this methodology is implemented in the case of Hellas. In this connection a revised stratification methodology was introduced in activity A in 1996. As a result, a new sample was obtained for the Hellenic regions of Macedonia and Thrace. This was based on limited number of strata and it used no satellite data. However, the methodology takes into account census elements such as the portion of the agricultural land at a local community or municipality level. The statistical evaluation of the new sample was obtained through the development of a geometric model between the observed coefficient of variation (CV) values and the percentage (%) of the land used for agricultural purposes. The model allows the calculation of the estimated CV values as well as the estimated precision of the area estimates. Finally, the third objective is to provide a comparison of the results obtained when both the Activities A and B are applied in Hellenic territory.

In Section 2, the main features of the methodology of the area frame sampling surveys are reviewed. The dimension that characterises an aerial sampling strategy from the statistical point of view is analysed and discussed in more detail. In Section 3, the area frame survey documents, such as the aerial photography and the satellite images, as well as the organisation of the work in the field are briefly described. Further, in Section 4, Hellenic experience in the framework of the activity A of the MARS-STAT project is reported. This experience includes the methodology adopted in order to obtain a suitable sample and the required steps needed to obtain the sample, using the current and the proposed stratification methodology, as well as the statistical evaluation of the results obtained. In Section 5, the area frame sampling methodology as it is implemented in EU in the framework of the activity B is presented. In this section, particular comparisons between the results obtained by implementing the Activities A and B in the Hellenic case are provided. Finally, in Section 6, the conclusions of this work are presented.

2. General features of area frame surveys

Council Regulation no 837/90 of EC is an obligation to all Member States of the EU, in order to provide annual information on areas cultivating main crops, such as wheat, maize and cereals. Since Member States use different systems for agricultural statistics, namely, village statistics, census, area sampling surveys, administrative by-products etc., the results obtained by the implemented surveys were based on statistical methods meeting the specific requirements of the agricultural system used in each Member State. From the statistical point of view this approach provides reliable and coherent methodology since the statistical services provided, and the statistical results obtained, rely upon different systems. However, as a result of the noticeable tendency towards the implementation of the RS and the GIS, the Council Regulation no 582/92 of EC came into effect, providing the framework for the implementation of the area frame sampling methodology, as part of the measures to be taken by the national governments to restructure their agriculture statistical surveys.

The main characteristic of an area frame survey is that the sampling universe is an area, for instance a region or a country, in which the sampling units are small parts of that area. A frame specifies the elements of a population out of which a sample can be drawn to estimate a certain characteristic of the complete population. When the population is finite, the frame may be defined by an explicit list of its elements. In agricultural statistics, this corresponds to sampling farms from a census supposed to contain all the farms of the region surveyed (list frame), or from an area frame of segments limited by physical elements of the landscape (cadastral segments). In the case of an infinite population, infinite sampling frames can often be considered as finite by allocating a size to each element. This would be the situation for agricultural surveys by point sampling in area frames, where a geographical point is taken as 1 m^2. However, on many occasions the frame is finite, but there is no need to build up an explicit list of their elements, as for example is the case when sampling square segments in an area frame.

An area frame survey is to be used over a period of 15–20 years. It is only one of the tools available to the statistician, who often has the choice between a list frame survey and an area frame survey, to obtain the information needed. It is an efficient tool to produce statistics on land use and yields. It can also prove useful when there is no farmer list available, for legal reasons (the United States), or for practical reasons (countries undergoing major changes in land tenure such as those of Eastern Europe). Area frame surveys can then be broken down into two categories: surveys on land use and yields and multi-purpose surveys. The first category provides information on acreage and yields only, while the latter provides information on all items relating to the farm. The above distinction is important when assessing the profitability of RS, in order to avoid the mistake of comparing RS, which provides information on acreages and yields only, to multi-purpose surveys which provide information on all items relating to the farm.

Surveys on land use and yields must be as simple, inexpensive, and easy to carry out as possible. In most of the cases, the joint use of remote sensing leads to a reduction in the size of the ground sample. The expected reduction of the ground survey cost has to compensate for the cost of the images and of their processing. In contrast, multi-purpose surveys are more complex since they are used to build the list of farmers which is not available. From this basis, it is possible to set up a system of surveys concerning all the aspects of farming (farm structures, yields, incomes, etc.). The following analyses in brief the dimensions that characterise an aerial sampling strategy.

2.1. Stratification versus non-stratification

Stratification is the division of a finite population into a number of non-overlapping sub-populations (strata) based on one or more characteristics. The closer the behaviour of the elements within each stratum the more efficient the stratification. Classically, the strata are defined so that each segment of the population belongs to one, and only one, of the non-overlapping strata. Therefore no element may be shared by two or more strata. In the case of an area frame made up of segments, this means that no segment straddles the border between two strata. An example is the division of a population on the basis of a quantitative variable (e.g. farms on the basis of the acreage). In that case, a list of the individuals concerned is needed prior to stratification. This list must mention the value of the variable for each individual. Usually, the latest survey on the subject provides the list.

If properly applied, a stratified sampling scheme gives a more precise estimate of the population parameter than a simple random sample of the same size (Cohran, 1977). The aim of the stratification is to reduce the sampling errors of the estimates, and to produce strata which are as homogenous as possible (in order to minimise variances within strata). In the case of area frame surveys, the individuals are area units, for instance square segments. In the case of a stratified survey the sampling strategy includes the choice of the stratification criteria, the choice of the number of strata, and the allocation of the sampling units to the strata.

A stratification is useful when the strata are as homogenous as possible in respect of the variable under study, or a variable correlated with it (i.e. a survey on wheat can be stratified on the basis of the percentage of arable land). It is often profitable to make use of the knowledge of agronomists to divide the territory on the basis of pedological, climatic and agronomical factors. However, one should take care not to split the area into very small homogeneous micro-zones, as it will then be necessary to merge them into operational sampling units. Note that although statistical theory indicates that strata must be as homogeneous as possible, even if this will result in an increase in the number of strata, it does not provide a method of computing the optimum number of strata before the survey.

Examples of stratification criteria used in some surveys may be found elsewhere (Cotter and Nealon, 1987). However, more refined stratification may be obtained with the use of multivariate algorithms (cluster analysis), by combining different layers of information in a more automatic way, or by improved use of satellite imagery (Cotter and Tomczac, 1994). During the first years of the MARS project the strata were drawn up from manually amalgamated available maps, statistical data for small administrative units, and in some cases, satellite images. The most common stratification tools are topographic or thematic maps, including land use maps, geological and pedological maps. Each stratum obtained is generally formed by one or a few relative polygons (continuous areas). The minimum size of a polygonon must not be smaller than the segment and must in principle represents at least ten segments. If statistical data are available for small geographical units, such as municipalities, a clustering procedure can lead to strata with a large number of scattered pieces. This is exemplified by the Hellenic case study, where there is a considerable amount of statistical information assimilated at municipality or local community level.

2.2. Nature of strata and sampling units

Field boundaries from the farmers' crop plants are digitised either from cadastral or topographic maps, or from aerial photos. In the case of cadastral maps the limits of the strata can be physical elements of the landscape such as roads, rivers, administrative boundaries, lines drawn on a topological document etc.. The quality of the boundaries is the most important concept, given the fact that the area frame is to be used over a period of 15–20 years, and that there is no difference between the nature of strata limits and the nature of last order sampling units (i.e. segments). However, the actual method used depends on the quality of the available documents and the type of referencing given by the farmer. The digitised field boundaries are kept as vectors but are used in tandem by superimposing them on raster satellite data.

Studies conducted within MARS project (Delince et al., 1993) have shown that cadastral maps, which are based on land ownership, are not in a good geometric agreement with actual cultural plot shapes. Automatic segmentation of Spot panchromatic and multi-spectral images has also been tested. The objective was to automatically detect individual field limits, and to provide farmers with documents containing the limits of their fields. Many algorithms were used without providing adequate recognition of the boundaries. Edge detection and region growing methods are very sensitive to threshold values, while clustering uses only spectral, not spatial, properties of the data, all leading to unstable results in the field boundaries. Furthermore, a vectorisation process was found inappropriate regarding area and shape accuracy, particularly when converting field delimitation from raster to vector (Terres and Arvain, 1993).

It should be noted that a stratum may be discontinuous and may have any form. Its size, however, must not be too small, especially if stratification is done not only to reduce variation in estimates, but also to report such estimates for every stratum of interest. If a regression analysis is used, then a minimum number of segments must be kept in every stratum. Thus, provided that the usual sampling rates of 0.5–1% lead to samples of 36–72 square segments of 50 ha in a Spot frame (360000 ha) it is clearly difficult to divide this frame into more than two strata.

There are three basic types of aerial sampling units that can be identified, namely, the point, the line and the field. We are particularly interested in the field type, which is based on the concept of a segment representing a closed land area. Note that a survey may concern acreage and yields only, or other farm data. Thus, in the first case, the survey is carried out on the area within the limits of the segment only (closed segment). In the second case, all farmers who own or use the land within the limits of the segment are interrogated (open segment). This is a way to compensate for the non-existence of a farmer list. Closed or open, the segment can be square, or irregular. The limits of square segments are lines drawn on the survey documents, which do not correspond to a physical reality. The limits of irregular segments are usually physical limits, as are the stratum limits. The major advantage of the square segment is that its sampling does not need a prior stratification, nor a preliminary knowledge of the territory. It is particularly adapted to Spot frames, whose limits vary as a function of the viewing angle, and cannot be known accurately before the acquisition of the images.

Advocates of the irregular segment claim that it is easier for the enumerator, namely the person doing the field work, to locate himself in the segment when the limits are physical. But good physical limits are rare, especially at the level of the segment, and when the limits are drainage and irrigation canals, intermittent streams, field boundaries, trails and internal roads, they are not very easy to find. It must be noted that, if the enumerator does use these physical elements as limits, there used as landmarks to locate the non-physical limits. Moreover, the search for suitable physical limits may introduce some bias into the sampling and therefore, this point has to be thoroughly examined. The Spanish Institute for Agricultural

Research (Instituto Nacional de Investigaciones Agrarias) has carried out a comparative study using both types of segments (Gonzalez et al., 1990). It was concluded that in both cases, the accuracy of the estimates obtained either by direct expansion or by regression are similar. Further, from the relative efficiency (the reduction of the variance due to the use of satellite data) point of view, it seems better to use square segments.

Theoretically, the optimal size of a segment is the one that gives the greatest accuracy for a given cost. Practically, this is difficult to determine. In order to reduce the survey costs, segments must be as small as possible, but the probability of finding physical elements (limits or landmarks) decreases with the size of the segment. Note, that if the segment is too small, it usually takes more time to travel between the segments than to survey each of them. In addition, the size of the segment should allow for its survey in one day or less, travel included. Exceptions may exist, but they are limited. However, the smaller the segment, the less representative it is of the global population, and the greater the variance between segments. The payment of the enumerators must be consistent with the remuneration of other surveys (e.g. face-to-face interviews). It depends on the time it takes to do the work, which in turn depends on the conditions of access (quality of the road network), the difficulty in finding the fields, the number of tracts, the complexity and the clarity of the nomenclature. Note, that a tract is defined as the set of fields or pieces of field inside a segment that belong to the same farm and that usually only the utilised agricultural area is considered in the tract.

If the sample is not stratified, or in the case of a stratification based on stable criteria, the area frame can be used over a period of years without having to update the sampling units. The problems arise from the necessity for regular updating of the ground survey equipment. The particularity of the equipment necessary to an area frame survey is that it necessarily includes a topographic document, generally an aerial photograph. This document does not always age well, particularly when regrouping of lands, land reclamation programs or major public works have been carried out. The photographs must be regularly updated, which may generate practical problems when transferring the outline of the segment or the cluster of points because of the different geometric distortion between the old and the new photographs.

3. Survey documents and work organisation

All types of area frame survey require the use of topographic material. This material helps the enumerator to find the sampling unit (road map at 1:200000 scale, topographic map at 1:50000, etc.), and to describe it (scale larger than 1:10000). These documents, especially the last-mentioned, must be introduced into a geographic co-ordinate system, in order to allow their location in the corresponding satellite images.

3.1. Aerial photography

The most utilised topographic document is the aerial photograph. This document has the advantage of being relatively cheap and having an accuracy which allows enlargements at a scale of up to 1:4000. The sampling unit can be drawn manually on the enlargement. In the case of a cluster of points, it is preferable to overlay a master grid to the negative of the photograph and make the enlargements from this document. The main technical drawbacks of the aerial photograph are its poor geometrical quality and the imprecision of its scale. The application of a standard enlargement factor to all the photos of the same area may lead to some error in the calculations of the acreage, and to a bad overlay of the segments on the satellite image. Corrections can be made, but the document is no longer as cheap in that case.

The use of aerial photography assumes the existence of a professional geographical institute or agency, able to update regularly the coverage of the country and to carry out some of the specific tasks required by the use of aerial photos as survey documents (geometric corrections, drawing of clusters of point, or segment limits, etc.), or to give the user the information necessary for these tasks. In the case when there is not such an availability of aerial photographs, it is necessary to set up an aerial photography mission, which can prove costly for an operational project. Note, that considerable modifications in the planning of the ground survey might be necessary if, for example, aerial photographs are not available and must be substituted by prints of satellite images or the use of the Global Positioning System (GPS).

3.2. Satellite image

Within the framework of the MARS-I project, JRC used satellite images as the basis of survey documents. This was made possible by the high resolution and repetitivity of Landsat-tm and Spot. The aerial coverage is not updated every year, but since satellites can usually take images of a given area approximately once a month, it is possible to have a new document every year. The quality of the document can be enhanced by digital processing of the images (Fierens and Rosin, 1994). Geometric corrections to make the survey document superimposable on a topographic map; filtering, contrast enhancement, and re-sampling to 5 m (which simulates a 5 m resolution) are some of the functions needed to be carried out.

The question of the location of the segment in the satellite image arises in different form from the case of aerial photography. Here the precise location of the segment in the image is known, and the enumerator must find it on the ground. There is no calculation to make, provided that the image is geometrically corrected. Nevertheless, the satellite image has the major drawback of not being able to be enlarged beyond 1:10000. Moreover, digital quality enhancements can generate unexpected artefacts that can puzzle the ground enumerator such as, checkerboard patterns, the widening of certain linear elements, the displacement of some limits. Thus, it is necessary to complement the satellite enlargement by a topographic document, which is, an 1:10000 enlargement of a topographical map. This is a major limitation for the countries which do not have a good aerial photographic coverage.

Some of the aspects favouring the use of satellite images as sole survey documents are large fields, not exceeding 30–40 per 50 ha segment; field boundaries running at right angles; prevailing land use for either annual crops or permanent crops, but not both in association; strong differentiation in cropping calendar between adjacent land cover types and finally the possibility to take into consideration the cropping when choosing the date of acquisition. For example, for crops in Europe, dates allowing the best discrimination between winter and summer crops are selected. A day shortly following the harvest of winter crops is in general the most favourable. However, if the interest is in irrigated crops only, dates will be earlier in the year. Features hampering the use of these images are land cover patterns with perennial and annual crops mixed, and irregular field shapes, particularly in scrub-land areas. Finally, documents from satellite images are comparatively expensive, and making a new set every year would increase the cost of the survey. Consequently, it is not always possible to take full advantage of the repetitiveness of the satellite.

3.3. Organisation of the field work

Depending on its aims, an area frame survey can be costly and sophisticated, or simple and adapted to situations where funds are limited. The elements necessary are the topographic documents to locate and delineate the sampling units as well as the enumerators able to use these documents.

The enumerator must locate the sampling units (segments, clusters of points) using a topographic and/or a road map and select the sequence in which the survey of the units allotted for the day will be made. In each sampling unit, the route will be planned before the beginning the survey, as a function of the location of roads, tracks, and obstacles such as streams or ditches. The survey of a cluster of points consists in determining the nature of the land cover for each point and writing it down on a questionnaire. The survey of a segment consists in plotting the track boundaries within each segment as accurately as possible on a transparent film, and noting the contents of each track.

To survey a segment, the field-worker will require the topographic documents (enlargement of aerial photograph or satellite image, enlargement of the topographic map, etc.), a questionnaire, a transparent film overlaid on the enlargement, and fine indelible markers for drawing on the transparent film. These necessary elements can be completed by a clipboard, a compass, a graduated ruler, a scale ruler, a 10-m tape, etc. The enumerator plots the tract boundaries (namely, the area of uniform land cover, whether the land is used for agriculture or not) and writes the number and the land cover for each tract on the questionnaire.

The nomenclature must meet the needs of the final user of the statistics, often a Ministry or a statistics institute, but it must not be a simple copy of the nomenclature used for official statistics. For instance, some crops like maize can be used for grain, silage, seed-production, game shelter, etc. Unless some external information is available, for instance from the farmer, the enumerator will not be able to classify the maize seen in one of these categories, but will note maize and nothing else. Therefore, as a general rule, the enumerator must not be asked to use code positions that cannot be discriminated in the field because it causes loss of time and confidence.

Although linear features like roads, railways or rivers rarely have a width of 20 m, they play an important role as topographic landmarks and tract limits. This is particularly true in the case of segments with topographic limits, but is equally true with a square in which this network of linear features is very important for the interpretation. Rules for drawing the segments depend on the required precision. The enumerators must be able to make a good drawing, but must simplify the observed reality in order to avoid complicating the drawing by irrelevant and useless details. The limits must be straight or broken lines, or simple curves. It is not necessary to plot all the irregularities and all the details. The supervisor must be vigilant, to avoid useless complication as well as oversimplification. Drawing rules also depend in the type of digitising hardware (digitiser or video camera). With a digitiser, the curves must be drawn as broken lines. With the video camera they may be drawn as curves. A close co-ordination between the ground survey team and the processing team is necessary for the good management of a project.

4. MARS-STAT activity A: regional inventories

4.1. The stratification methodology in Hellas

For several years a number of regions in the EU carried out regional inventories using the methods developed by the MARS project. The project assured the follow-up of these activities and kept on investigating possible enhancements of the methods applied. The transfer of the methodology to the regions, which was seen as the major objective, was almost completed in 1994. From 1993 onwards a gradual geographic shift of the MARS project support activities took place towards the central European countries and certain regions in southern Europe. In addition, the project continued to supply technical support to the regions in the EU and to intensify its collaboration and training activities to the PHARE countries.

The methodology adopted establishes close links between satellite data and observations on the ground. Development and evaluation focus on the so called regression estimator method. The action comprises two components, the first of which aims to obtain objective observations in the field with a sample design established or enhanced by RS, while the second introduces automatic classification of the satellite data techniques in order to improve the regression-based estimates generated by the ground surveys. The latter component involves stratification using satellite images; the selection of a sample ground survey on the basis of aerial photographs, using existing topographical documents and maps; the simultaneous acquisition of a full coverage of the region by Spot and Landsat-tm images; the automatic classification of the satellite data in order to improve the regression based estimates generated by the above mentioned ground surveys and, finally, the analysis of the results. RS also comes into play by providing the enumerator with documents enabling the accurate plot on the ground and the correct location of the sampling units.

In Hellas, a stratified sample of the most important cultivated areas, consisting of 2043 sampling units or segments, was gradually obtained during the years 1989–1994. The distribution of segments in the sample is shown in Table 1 (NUTS 2 level). However, for comparison reasons with the new sample obtained in 1996, the pilot regions of Macedonia and Thrace are presented in more detail (NUTS 3 level). Note that regions in Europe are classified by the SOEC into Nomenclature of Statistical Territorial Units, or the so called NUTS levels. Thus, NUTS 0 refers to the countries, NUTS 1 refers to standard statistical regions, NUTS 2, scaled approximately, refers to large European regions, such as the ‘province’ in France and Belgium, the ‘regierungsbezirk’ in Germany, and the ‘region’ in Hellas (An-

Table 1
Distribution of segments in the current and the proposed sample

Region (Nomos)	NUTS 2	NUTS 3	No of segments	
			Old sample	New sample
East Macedonia and Thrace	RA11		331	268
Evros		RA111	111	99
Xanthi		RA112	38	27
Rodopi		RA113	57	57
Drama		RA114	88	47
Kavala		RA115	37	38
Central Macedonia	RA12		434	442
Imathia		RA121	40	36
Salonika		RA122	105	93
Kilikis		RA123	60	61
Pella		RA124	53	54
Pieria		RA125	33	32
Serres		RA126	80	100
Halkidiki		RA127	63	66
West Maceodonia	RA13		174	167
Grevena		RA131	41	37
Kastoria		RA132	29	25
Kozani		RA133	67	72
Florina		RA134	36	33
Thessalia	RA14		233	
Epirus	RA21		101	
West Hellas	RA23		165	
Sterea Hellas	RA24		194	
Peloponnisos	RA25		215	
Crete	RA43		196	
Total			2043	877

dalusia in Spain, Lombardy in Italy, Flanders in Belgium and Thessalia in Hellas are some typical examples), while NUTS 3 correspond to the smaller administrative units such as the ‘department’ in France, the ‘kreis’ in Germany, the ‘district’ in Belgium and the ‘nomos’ in Hellas.

The first phase of the implementation (1989–1991) started from Central and West Macedonia (RA12, RA13) and ended with East Macedonia and Thrace (RA11), giving a sample of 939 segments (Hellenic Ministry of Agriculture, 1991). During the second phase (1992), Thessalia (RA14) and some parts of Sterea Hellas (RA24; RA243, RA244) were included in the sample, increasing the total number of segments to 1280 (Hellenic Ministry of Agriculture, 1992). In the third phase (1993) the remaining parts of Sterea Hellas (RA24; RA241, RA242, RA245), and the regions of West Hellas (RA23) and Peloponisos (RA25) were included in the sample, increasing the total number of segments to 1748 (Hellenic Ministry of Agriculture, 1993). Finally, in the last phase (1994), the regions of Crete (RA43) and Epirus (RA21) were included in the sample, giving the total number of 2043 segments (Hellenic Ministry of Agriculture, 1994). Since 1993 all the work was carried out by the Ministry in collaboration with the InfoLab of the Agricultural University of Athens, while for the previous years the stratification and the sample selection was carried out by the JRC in close co-operation with the Ministry and private subcontracting companies.

The strata and the sampling units were obtained using statistical data about the cultivated areas, Landasat-tm satellite images and topographic maps. In particular, the statistical data and the crop calendar were used in order to locate the cultivated areas and to find the appropriate satellite scenes, while the 180×180 km with 30×30 m resolution Landasat-tm satellite images were used in the identification of seasonal cultivation. For the Hellenic regions examined, the useful dates to identify cereal crops were between May and June, while for summer crops, irrigated or non-irrigated, it was better to choose between July and August when they were at the peak of chlorophyll activity and the spectral response was the highest in the near infra-red. The dates allowing the best discrimination between summer and winter crops were at the beginning of July. Only bands 2 (0.52–0.60 μm), 4 (0.76–0.90 μm) and 5 (1.55–1.75 μm) were actually needed. Note, that spectral band 2 has good response in the absorption of chlorophyll activity in the region between the blue and red and corresponds to the reflected green of the vegetation. Band 4 has very good response in the existing biomass of an image and it helps in the identification of crops. Finally, band 5 is sensitive in the amount of water that is contained in the crops. The images were used in ‘false colour’ according to the common conventions; red, green and blue correspond to 4, 5 and 2, respectively.

Topographic maps of 1:50 000 and 1:100 000 scale were used in order to correct the geometry of the images, georeference hard copies, rectify the geometry of ARC/INFO coverages created from images that were not geometrically corrected, help in the photo-interpretation and determine the NUTS 2 and NUTS 3 boundaries. The geometrical correction of the scenes was made with the help of the control points that were easily identifiable in the scenes (scene co-ordinates) and the topographic maps (geographical co-ordinates). Then, the administrative boundaries of the region under consideration (usually NUTS 2 level) were placed on the scene. Note that in the case of geometric correction of images the expected error with 1:100000 scale maps was found to be around 2 pixels, or 60 m on the ground and therefore, for the required degree of accuracy for stratification (precise geographical position of the segments), it was not possible to use maps at lower scale. All the above work was carried out either on a PC or a VAX workstation using ERDAS and ARC-INFO/ARC-VIEW software (Hellenic Ministry of Agriculture, 1991–1994).

The characteristics and the number of the strata depended upon the agromorphological characteristics of the region. Seven strata were defined, namely, high mountains, mountains and hills with crops, hills and plains of the region of Halkidiki, high plains and basins, non-irrigated plains and hills, irrigated plains, and the Axios Delta. The definition of strata was mainly based on the photo-interpretation of the radiometric responses. Note that images are usually obtained at the same date on the same path, so that it is possible to create a mosaic. Therefore, at a given date, crops in full chlorophyll activity, when the spectral response was the highest in the near infra-red, corresponded to a well defined crop type which constituted a stratum.

However, for certain strata, the relative intensity of the colours was not clearly distinct. The problem was overcome by looking at the texture of the image. It was noticed, for example, that a stratum of permanent crops-polyculture was identified by its texture, which differed from the linear aspect of the plots of annual crops. It was also distinguished from other strata by the location of the permanent crops mainly on the slopes and by the good representation of the vines, orchards, olive trees on the topographic maps. Intensively irrigated crops or cereals on hills were often seen as homogeneous areas. On the other hand, a stratum of mixed crops (irrigated or non-irrigated summer crops, winder crops, permanent crops) was heterogeneous radiometrically but well identified by other characteristics as, for example, the size of fields and the location of the stratum in plains.

To obtain the sample of segments, each NUTS 2 region was divided into square blocks each of which was allocated an area of 14×14 km. Thus, with a segment size of 700×700 m (or 49 ha), a total of 400 segments was produced in each block. The stratification method used was the systematic aligned sampling method with a distance threshold of 3.5 km (Gallego, 1995). This means that the replications of segments in each block and from block to block have a distance of at least 3.5 km. The sampling pattern of replications (see Fig. 1) was kept the same in each block with as many segments as defined by the sampling rate. This rate was varied between 0.3% (mountains, forests, etc.) to 2.5% (plains with intensive cultivation). As a result, the number of replications in each block varied between three and seven. The selection was made at random and the sampling pattern was repeated across the survey region giving the 2043 segments which correspond to an area of 100107 ha. The number of replications in each block and their location in Cartesian co-ordinates were derived using a special software program. These co-ordinates were transformed into geographical co-ordinates and the corresponding segments were located on the topographic maps. Thus, the aerial photographs and their

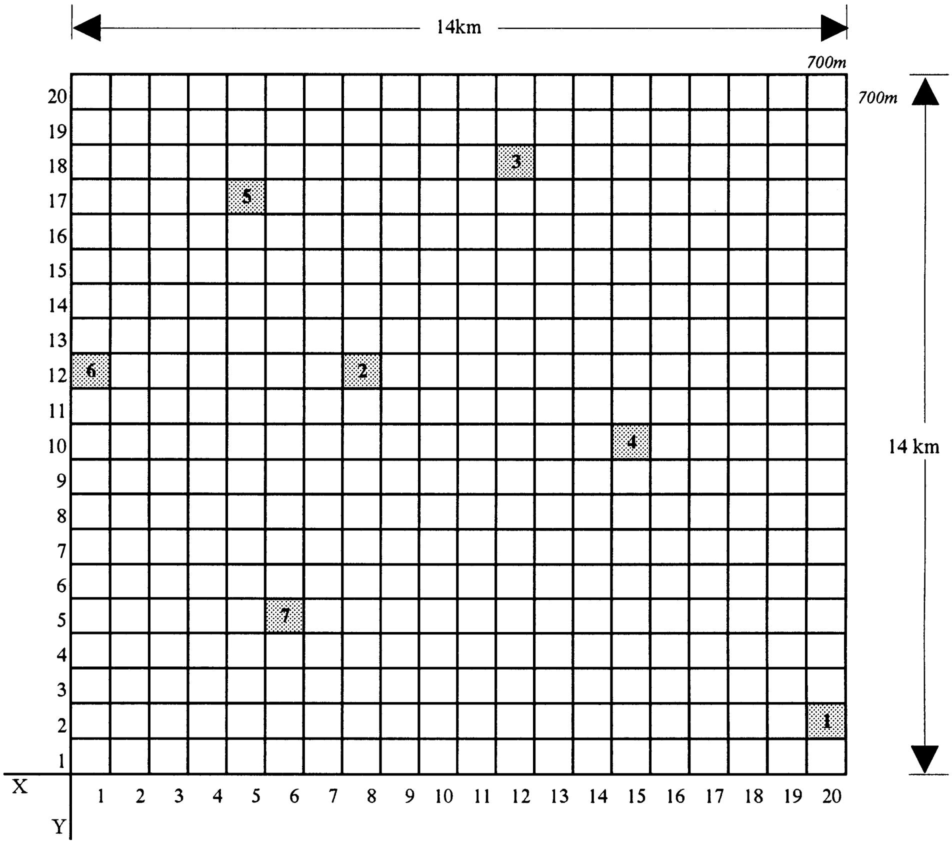

Fig. 1. Sampling pattern of replications.

enlargements (1:5000) were produced. Note, that the map scales used were 1:200000 for general overview and 1:50000 for locating the segments. Fig. 2 shows the stratification map of the NUTS 2 region of Central Macedonia (RA12) with their NUTS 3 sub-regions.

The enumeration in the sample segments was made by a group of agronomists and topographers, usually during the period of May through June, when the winter cereals and the summer crops were in the fields. The material used to enumerate the area frame segments were the 1:50000 maps showing the position of each segment in the sample, the 1:5000 aerial photographic or topographic map for each segment, a questionnaire and some A4 transparencies. Using the maps and the aerial photographs the topographer identified the location of each segment on the ground, and traced on the transparency the boundaries of each segment and individual plot. The agronomist determined the land use by completing a special questionnaire. Note, that during this stage a sample of farmers was obtained from a preselected number of segments (at about 30% of the total number of segments), with the aim to obtain yield estimates through a special questionnaire.

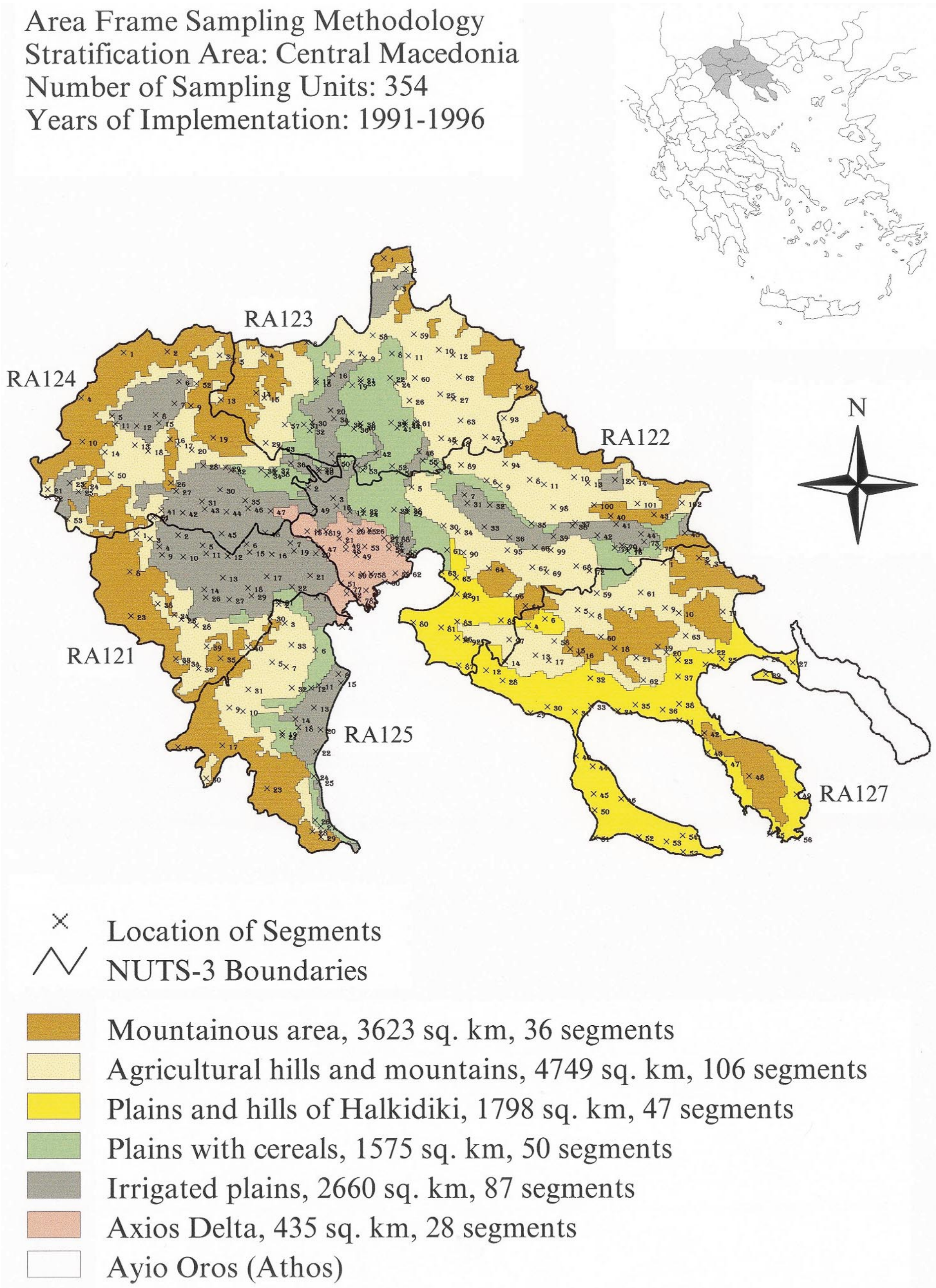

Fig. 2. Digitised stratification map of Central Macedonia, Hellas (current sample).

Table 2
Results of the estimated areas (ha) and CV in % during 1991–1996

No	Crop	1991		1992		1993		1994		1995		1996	
		Area	CV	Area	CV	Area	CV	Area	CV	Area	CV	Area	CV
1	Soft wheat	259 473	6.57	352 467	5.47	339 734	5.28	276 448	5.56	251 658	6.25	216 705	6.54
2	Durum wheat	398 298	5.31	510 000	4.47	498 497	4.11	559 701	4.00	558 225	3.94	549 410	3.96
3	Barley	91 712	7.85	126 751	6.96	135 138	7.73	136 808	6.34	111 989	6.66	109 431	8.10
4	Maize	82 817	12.18	138 305	7.03	176 468	6.00	191 882	5.52	138 759	6.13	246 571	6.14
5	Cereals					1 176 616	2.65	1 204 636	2.66	1 098 985	2.70	1 158 595	2.76
6	Cotton	67 479	22.08	318 338	4.90	365 783	4.54	395 446	4.31	466 713	3.98	450 021	4.08
7	Tobacco	25 405	13.84	61 681	11.75	71 906	9.96	62 469	11.20	52 303	10.71	47 080	10.69
8	Sugar beets	37 493	8.36	62 855	10.18	57 526	10.23	45 547	9.84	44 969	10.00	60 309	11.46
	Cultivated area	4 241 980		6 384 880		9 884 580		11 638 500		11 638 500		11 525 890	

Table 3
CV (%) comparison between the area frame sampling and the image analysis methodologies

Crop	CV (%) results					
	1992 RA12, RA13 (3 384 200 ha)		1993 RA14, RA23 RA24, RA 25 (5 642 600 ha)		1994 RA14 (1 403 660 ha)	
	Area frame	Image analysis	Area frame	Image analysis	Area frame	Image analysis
Soft wheat	7.3	7.0	13.9	11.9	26.1	24.2
Durum wheat	6.6	4.9	5.9	4.4	9.4	7.6
Barley	9.4	7.8	16.3	12.9	18.6	14.7
Maize	11.1	5.1	8.9	8.5	16.3	14.9
Cotton	11.6	4.9	5.2	3.7	5.8	3.4
Tobacco	13.5	12.9	17.9	16.5	43.9	39.7
Sugar beets	17.0	5.6	20.9	13.1	25.8	13.4

All the above transparencies were digitised and the resultant files were processed to obtain the initial results regarding the area estimates (Table 2). The digitisation was made on a PC using a frame grabber and a special video camera (512×512 pixels) with the television tracking system (TTS) software, specifically developed by the JRC (Annoni and Gallego, 1992; Hellenic Ministry of Agriculture, 1991–1994). The effective resolution in the output was 360 pixels, equivalent to, approximately, 2 m for segments of 700×700 m. Thus, although the resolution of the system was not sufficient to efficiently digitise strata limits or administrative units, it was sufficient to digitise segments for statistical purposes. Finally, the desired area estimates were calculated using a special software program, that directly processes the produced *tts* files.

The above results were compared with those obtained using the unsupervised classification methodology (Table 3). The required satellite image analysis was made using the files that had been produced from the field based digitisation. The image processing was made on a VAX workstation, using the MARS-PED software (Annoni and Gallego, 1992; Hellenic Ministry of Agriculture, 1991–1994). Details of the method itself may be found in Hill and Aifandopoulou (1990). Note, that each year the required satellite image analysis was carried out in different regions of the Hellenic territory (Hellenic Ministry of Agriculture, 1994). The comparison shows the accuracy of the results obtained from the image analysis method. However, as may be observed, there is no great disparity between the two methods in most of the cases considered. Exceptions exist in the cases of maize, cotton and sugar beets in RA12 and RA13, for the year 1992 and perhaps for sugar beets in RA 14 for the year 1994. In these cases the results of the image analysis appear to be superior. A possible explanation is because the specific images were acquired at an appropriate time regarding the development of these particular crops.

4.2. The revised stratification methodology

Preliminary studies conducted by the Hellenic Ministry of Agriculture and the Agricultural University of Athens, showed that in some cases the characteristics of the strata used were inconsistent from one region to another (Hellenic Ministry of Agriculture, 1996). In addition, in many cases, the number of strata used was unnecessarily large and as a consequence, the cost of the survey was very high. For example, it was noted that in some regions interest was given to physical and geographical characteristics of the landscape, while in some other regions interest was given to crop intensive elements, such as irrigated or non-irrigated areas. However, the interest was clearly on crop and crop tree areas, their acreage estimates, and finally, in obtaining the sample of producers who would be used in order to find the estimates on yield production.

In 1996 a new stratification methodology was implemented and a new sample of the regions of Macedonia and Thrace (RA11, RA12 and RA13) was obtained. The new stratification approach was based on information provided by the National Statistical Service of Greece (NSSG), namely, the portion of the land in each municipality or local community that was used for agricultural purposes (agricultural land). Note that agricultural land includes crop and crop tree areas, arable land and fallow land. Thus, ignoring waters (lakes, rivers, etc.) the agricultural land of any local community or municipality was divided into four percentage classes, each representing a stratum, i.e.: less than 15%, between 15 and 50%, between 50 and 75%, and finally greater than 75%. The above classification was applied on the pilot area and the results are presented on the digitised map shown in the Fig. 3. The method used was the same, namely, systematic aligned sampling with a distance threshold of 3.5 km and up to seven replications on each block. Thus, the segment size was taken again 700×700 m, while the replications in the strata 1, 2, 3 and 4 were up to 1, 5, 6 and 7, respectively. Finally, the distribution of the segments in the final sample (NUTS 3 level) is presented in Table 1, while their locations are shown on the digitised map in Fig. 4.

4.3. Evaluation

The precision of any survey is related to the variability of the individuals in the surveyed population, to the sample size, to the size of the population which in this case corresponds to the total area of the region, and to the efficiency of the stratification. In order to see the likely precision that could be obtained for a particular crop in a region, a simple regression model was applied, using the complete set of the area estimates reported in Table 4. The data (1996) suggest a geometric relationship of the form, $y\sqrt{n} = \beta x^{a}$, where y presents the observed CV values (in %) of the crop area estimates, n is the number of segments in the sample and x (in %) is the estimated crop area. The geometric model above may become

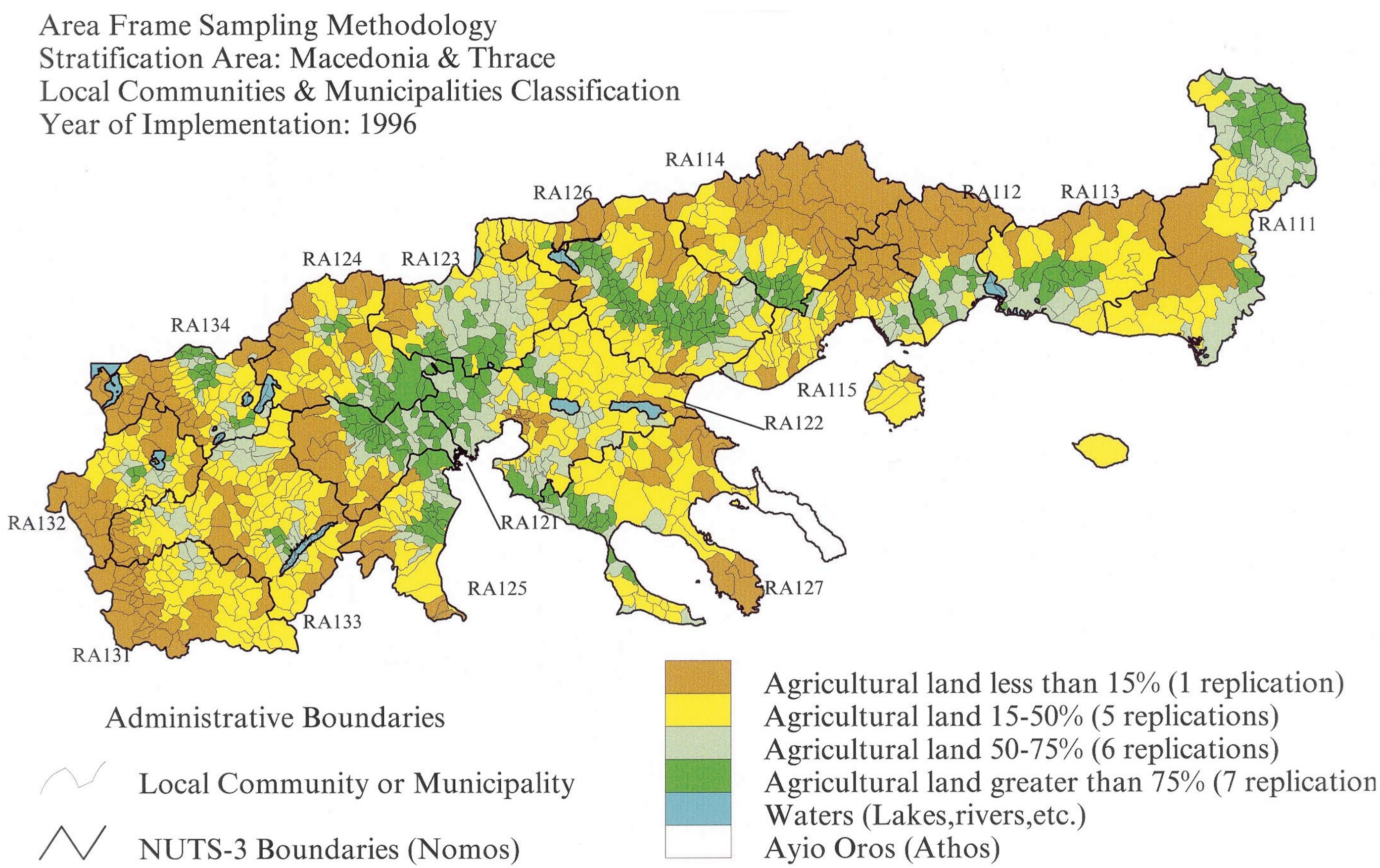

Fig. 3. Digitised stratification map of the Hellenic regions of Macedonia and Thrace, based on local communities and municipalities classification (revised methodology).

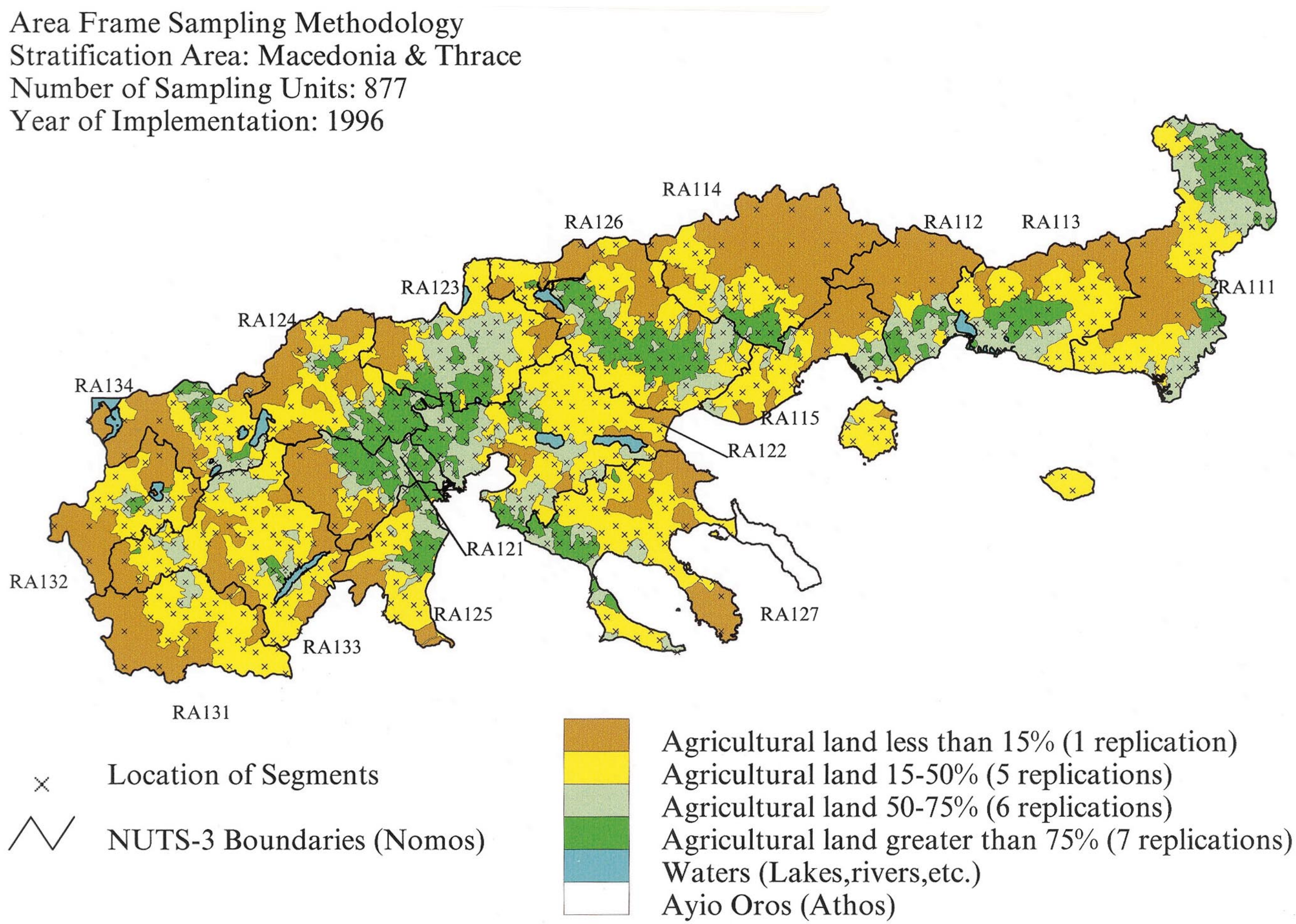

Fig. 4. Digitised stratification map of the Hellenic regions of Macedonia and Thrace (proposed sample).

linear straightforwardly, giving $Y = AX + B$, where $X = \ln(x)$, $Y = \ln(y\sqrt{n})$, $A = a$ and $B = \ln(\beta)$. The estimates of the parameters resulting from the above modelling are given by $\hat{a} = -0.3391$, and $\hat{\beta} = \exp(5.5706) = 262.58$, with the determination coefficient $r = -0.9551$. Therefore the model may be rewritten as $262.58x^{-0.3391}/\sqrt{877}$ and it will be called model (1). Since the test H_0: $r = 0$ against H_1: $r < 0$ suggests to reject the H_0 at $^*P < 0.05$ we conclude that the determination coefficient of the corresponding linear model shows a very high negative relationship between the examined ln variables x and y. This result is presented in Fig. 5.

The results obtained from the above modelling are compared with those obtained from the model (2): $\hat{z} = 371.5x^{-0.462}/\sqrt{877}$ suggested by Gallego (1995) and obtained from ground surveys by segments in several regions in Italy, Belgium, Greece, Portugal and Spain for different crops between 1988 and 1992. These results appear in Table 4. Further, in Table 5, the cultivated areas (in ha and %) per stratum of the most important crops in the regions of Macedonia and Thrace are presented. This information was recorded in 1991, during the last national census, and was provided by the NSSG. The above percentages were fed into both models

Table 4
Observed and estimated coefficient of variation (year: 1996) model (1): $\hat{y} = 262.58x^{-0.3391}/\sqrt{877}$ and model (2): $\hat{z} = 371.5x^{-0.4621}/\sqrt{877}$

No	Cultivation	Area		CV		
		β (ha)	x (%)	y (%)	$\hat{y}$ (%)	$\hat{z}$ (%)
1	Soft wheat	216 705	1.880	6.54	7.16	9.37
2	Durum wheat	549 411	4.767	3.96	5.22	6.10
3	Barley	109 431	0.949	8.10	9.03	12.85
4	Rye	20 232	0.176	29.09	15.98	27.99
5	Oats	36 477	0.316	16.76	13.10	21.36
6	Maize	246 471	2.138	6.14	6.85	8.83
7	Rice	33 520	0.291	15.42	13.48	22.19
8	Peas	176	0.002	56.39	72.94	221.50
9	Beans	3369	0.029	27.04	29.46	64.39
10	Broad beans and lathyrus	451	0.004	46.07	57.66	160.81
11	Other beans	4720	0.041	33.59	26.19	54.87
12	Potatoes	20 829	0.181	22.20	15.83	27.63
13	Sugar beets	47 080	0.408	10.69	12.02	18.96
14	Sunflowers	21 741	0.189	17.92	15.60	27.09
15	Cotton	450 021	3.904	4.08	5.59	6.69
16	Tobacco	60 310	0.523	11.46	11.05	16.92
17	Other industrial crops	1770	0.015	34.06	36.83	87.32
18	Fresh vegetables	41 238	0.358	9.33	12.56	20.16
19	Fresh fruits	144 532	1.254	8.44	8.21	11.30
20	Vineyards	112 296	0.974	9.58	8.95	12.70
21	Olive trees	748 032	6.490	5.31	4.70	5.29
	Cultivated area	2 868 812	24.089			
	Total area	11 525 890	100.00			

Table 5

Estimated precision of area estimates using the new stratification sample of the regions of Macedonia and Thrace (RA11, RA12 and RA13). Model (1): $\hat{y} = 262.587x^{-0.3391}/\sqrt{877}$ and model (2): $\hat{z} = 371.5x^{-0.4621}/\sqrt{877}$

Crop	Total		Stratum 1		Stratum 2		Stratum 3		Stratum 4	
	Area in ha |sd| in ha (and in %)		Area in ha (and %) CV in % (and |sd| in ha)		Area in ha (and %) CV in % (and |sd| in ha)		Area in ha (and %) CV in % (and |sd| in ha)		Area in ha (and %) CV in % (and |sd| in ha)	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Soft wheat	229 919.6		15 924.6 (1.1734)		109 655.7 (6.5376)		64 833.3 (10.4334)		39 506.0 (7.5321)	
	6172.9	6942.9	8.40	11.65	4.69	5.27	4.00	4.25	4.47	4.93
	(2.68)	(3.02)	(1337.5)	(1855.4)	(5143.9)	(5777.8)	(2595.5)	(2752.6)	(1766.3)	(1949.8)
Durum wheat	424 888.0		127 897.4 (7.6252)		146 314.0 (23.5459)		139 875.4 (26.6683)			
	8357.5	7515.3	9.58	13.94	4.45	4.91	3.04	2.91	2.91	2.75
	(1.97)	(1.77)	(1034.8)	(1505.7)	(5694.6)	(6276.5)	(4444.7)	(4265.0)	(4073.4)	(3849.4)
Maize	147 831.1		3042.2 (0.2242)		28 500.4 (1.6992)		38 655.9 (6.2208)		77 632.6 (14.8026)	
	3959.6	4540.9	14.72	25.03	7.41	9.82	4.77	5.39	3.56	3.61
	(2.68)	(3.07)	(447.9)	(761.6)	(2111.3)	(2798.6)	(1844.1)	(2084.1)	(2760.3)	(2804.3)
Cereals	765 033.7		40 812.8 (3.0073)		295 387.2 (17.6111)		231 715.5 (37.2893)		197 118.2 (37.5821)	
	12 908.9	12 5580.0	6.10	7.54	3.35	3.33	2.60	2.36	2.60	2.35
	(1.69)	(1.64)	(2491.2)	(3078.5)	(9901.9)	(9846.9)	(6022.9)	(5461.9)	(5110.0)	(4629.6)
Tobacco	43 460.8		5 296.8 (0.39)		20 192.7 (1.20)		8367.5 (1.25)		9603.8 (1.83)	
	2043.3	2933.7	12.20	19.37	8.33	11.15	8.02 (670.7)	10.93	7.22 (693.6)	9.49 (911.0)
	(4.70)	(6.75)	(646.2)	(1026.2)	(1681.3)	(2325.0)		(914.8)		
Cotton	67 371.7		183.1 (0.0135)		5179.9 (0.3088)		20 663.7 (3.3253)		41 345.0 (7.8827)	
	2296	2736	38.18 (69.9)	91.7 (167.9)	13.21	21.59	5.90	7.20	4.40	4.83
	(3.41)	(4.06)			(684.1)	(1118.2)	(1219.1)	(1487.9)	(1820.2)	(1998.1)
Cultivated area	9 180 300		1 357 100		1 677 300		621 400		524 500	

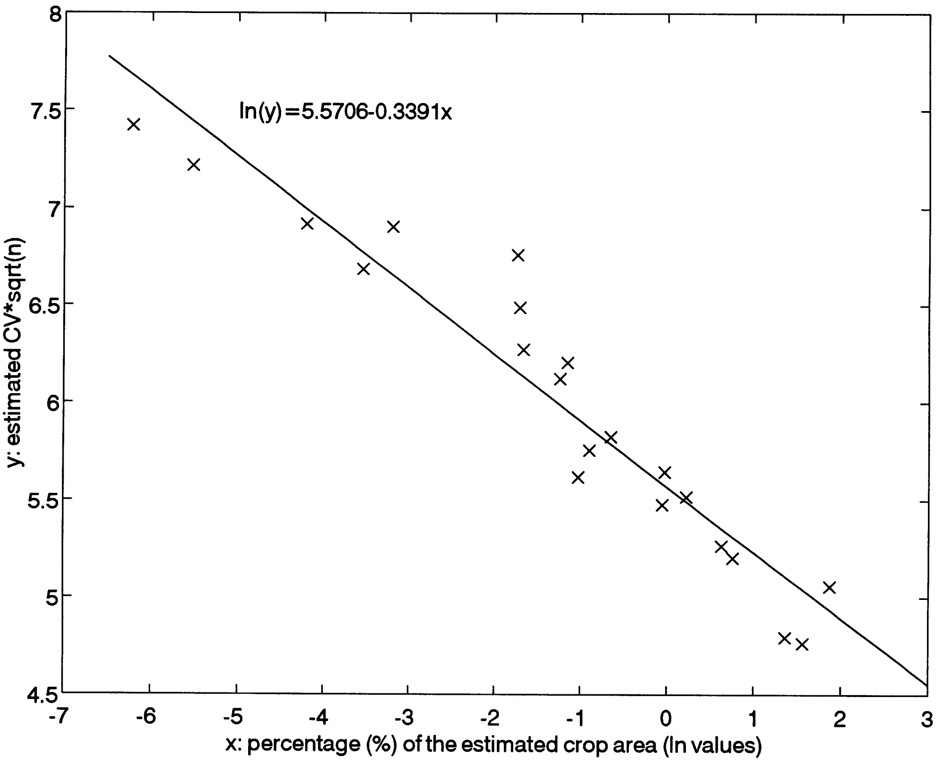

Fig. 5. Regression model.

(x values) to produce the estimated CVs (%) and the absolute standard deviations (in ha) per crop, per strata for both models. It can be seen that the absolute standard deviations obtained by the model (1) are lower and more accurate than those produced by the model (2). However, both models are proven satisfactory in terms of the total absolute standard error derived. For example, one may observe that in the case of cotton, although the absolute standard error in stratum 1 is 69.9 ha for model (1) and 167.9 ha for model (2), when the corresponding estimated total area is 183.1 ha, the total absolute standard error (i.e. considering all four strata) is 2296.1 ha compared with the estimated total area of 67371.7 ha (or 3.41%) for model (1) and 2735.8 ha compared with the estimated total area of 67371.7 ha (or 4.06%) for model (2). This is reasonable because area frame sampling is a methodology appropriate to estimate large crop areas. In the above case, cotton is very limited in stratum 1 and therefore the expected standard errors in this stratum are large for both models.

Finally, it should be noted that one may produce similar estimates for the absolute standard deviations taking into account the sample size per strata. Further research is currently going on, using satellite data, with the aim to investigate if there can be any further improvement on the above results.

5. MARS-STAT activity B: rapid estimates

5.1. The stratification methodology in the EU

Rapid estimates of the crop area are being produced in the EU by analysis of high resolution Spot and Landsat-tm satellite images on a sample of sites. Estimates are performed for inter-annual variations in the cultivation of the most important crops (Sharman and De Boissezon, 1991; Pous et al., 1995). The information in the images is also used to provide estimates of the potential yield of these crops. The results, produced throughout the crop year, are transmitted to the DG-VI of the EU and to SOEC, where they help to provide precise and up-to-date information on agricultural production. The method is based on a non-random sample of 53 square sites of 40×40 km. With the inclusion of three new member states (Austria, Finland and Sweden), the number of sites has become 60. Image classification for each site is performed on 1–4 high resolution satellite images per site during the agricultural season (March–October). Ground truth for training the classifier is obtained by photo-interpretation of approximately 16 segments 700 m or 1.4 km per side. Segments are selected at random by stratified sampling or located on a regular grid if there is no stratification. Image analysis requires a good knowledge base built on ground observations. Information comes mainly from ground surveys on the same sites in preceding years. Note that two types of ground surveys are usually conducted on the same sites; visits on the ground to a sample of points inside each segment, and a farm survey on an area frame sample (Carfagna and Gallego, 1995).

Central to the procedure is the so called conjuncturalist, who is responsible for the final estimates, and who combines as much information as possible to monitor the image analysis results. The quality of the final results depends on the ability of the conjuncturalist to combine different sources of auxiliary information with the image analysis. The conjuncturalist is responsible for the acceptance of classification results, as well as for the extrapolation of results, using data from previous years, to distribute an image classification group (cereals for example) into single crops. Auxiliary information drives the image analysis process, giving indications on the results that are likely to be true. Note that, no Bayesian methodology has been applied so far.

The method adopted has a number of advantages. First, it allows the delivery of a regular report with results updated on the basis of recent information. Second, the introduction of ancillary knowledge by the conjuncturalist allows to obtain better precision than the minimum possible standard error if the estimates were only based on the sample sites. Finally, the procedure might be adapted to get area estimates outside the EU, although it requires a significant amount of auxiliary information, that is more difficult to obtain outside the EU.

However, a number of limitations also exist. The method assumes a parallel evolution of crops that cannot be distinguished in the images. If for example farmers start to grow durum wheat instead of soft wheat, barley instead of wheat, or one summer crop instead of another one, this will not be detected. In addition, no statistical yield estimation is obtained from image analysis, although the conjuncturalist can give subjective estimates. Further, the sampling of sites is non-random, although ongoing studies improve the extrapolation of estimates. Finally, estimates are based on pixel counting, which is known to be strongly biased (Card, 1982; Hay, 1988; Czaplewski and Catts, 1992).

The sample of 53 sites used for rapid estimates in the EU (12 countries) was originally selected using stratification methodology and taking into account the satellite orbits and country borders. A number of sites were later substituted because they seemed to give little information on agriculture. Altogether the procedure has a strong subjective component and consequently extrapolation of the results presented considerable problems. A more objective approach has been followed to draw a complementary sample of seven sites in the three new member states, as well as for the new sample sites in Central Europe (PHARE countries) in the frame of the MERA project, initiated in 1995. The two-phase sampling procedure, with a first systematic phase and a further sub-sampling with a probability proportional to an index of agricultural intensity, is described in Vossen et al. (1995). The approach is based on square sites of 40×40 km and it was suggested by the size and shape of currently marked images, approximately square. This shape of sites is likely to give much redundant information for statistical purpose because of a high spatial correlation at relatively short distances.

An additional topic to be studied in more detail for different two-step sampling plans (sites and segments) is the accuracy improvement that can be achieved by obtaining exhaustive information on the first-step sampling units (sites) instead of observing only a sample of segments. However, ongoing research investigates the possibility of improving the precision of the estimates using sampling strips, namely sites with a long and thin shape, since the within-site correlation and hence the redundant information becomes lower. Note, that the accuracy of the results obtained by the rapid estimates activity throughout the agricultural season is evaluated by comparison with the results obtained using reference statistics published at the end of the year by the Eurostat. Nevertheless, the timeliness of the information and especially the early dates for which already stable and rather precise estimates could be obtained, when compared to other sources of information, contributes significantly to the satisfaction of the end-user. The lack of prediction data at national level is inherent in the sampling method. For countries that have ten or so sites, the prediction for two or three crops appears to be roughly valid. Such information at sub-European level, however, must be viewed with caution, as the sample is not designed for this purpose.

5.2. The Hellenic case study

By the end of 1996 the area rapid estimates campaign integrated for the first time seven new sites in the three new member states. The monitoring of cultivated areas is based on a survey using a statistical sample of 60 sites, covered by satellite images once, twice, thrice, or four times per year. A total of 171 out of 183 images programmed from the beginning of the campaign have been acquired. Taking into

Table 6
Comparison table between the area, yield and production estimates of the main cultivations in Hellas, during 1995–1996 obtained from the MARS Activities A and B

Crop	Area (in kha)				Yields (t/ha)				Production (in kt)			
	1995		1996		1995		1996		1995		1995	
	(1)	(2)	(1)	(2)	(3)	(4)	(3)	(4)	(5)	(6)	(5)	(6)
Soft wheat	251.7	252.3	216.7	260.0	2.90	2.84	2.32	2.60	714	717	503	675
Durum wheat	558.2	605.0	549.4	613.0	2.60	2.40	2.15	2.20	1445	1422	1181	1350
Barley	112.0	133.0	109.4	139.0	2.90	2.90	2.28	2.70	323	384	250	376
Maize	138.8	160.0	246.6	164.0	9.80	9.80	9.48	9.90	1358	1566	2338	1617
Sugar beets	45.0	42.0	47.1	46.0	71.00	60.40	64.49	61.00	3176	2561	3036	2811
Sunflowers		22.0		23.0		1.40		1.30		31		30
Potatoes		52.0		41.0		19.50		20.50		1006		844
Oats	38.4		36.5		0.200		1.71		77		62	
Cotton	466.7		450.0		0.300		2.33		1409		1049	
Tobacco	52.3		60.3		0.220		2.31		116		139	
Vineyards	105.7		112.3									
Olive trees	689.5		748.0									
Other crops	54.9		1.8									

(1) Area estimates based on MARS-STAT: activity A.
(2) Area estimates based on MARS-STAT: activity B (2 sites: RA12/Thessaloniki and RA14/Farsala, three images per site, per year).
(3) Yield estimates based on MARS-STAT: activity A.
(4) Yield estimates based on MARS-STAT: activity B (experimental, based on agro-meterological model outputs and Avhrr indicators at NUTS 0 level in combination with time series analysis).
(5) Production estimates based on MARS-STAT: activity A.
(6) Production estimates based on MARS-STAT: activity B (experimental).

Table 7
Wilcoxon rank-sum test for comparison of the population means between the area estimates of the main cultivations in Hellas, 1994–1996, obtained from the MARS Activities A and B

Year	x	y	x	y
	Soft wheat ($m=3$, $n=3$)		Durum wheat ($m=3$, $n=3$)	
1994	276.5 (5)	277.0 (6)	559.7 (3)	594.0 (4)
1995	251.7 (2)	252.3 (3)	558.2 (2)	605.0 (5)
1996	216.7 (1)	260.0 (4)	549.4 (1)	613.0 (6)
	$w_x=8$, $w_y=13$, $u_x=2$, $u_y=7$		$w_x=6$, $w_y=15$, $u_x=2$, $u_y=0$	
	$2\,\Pr(U \le 2 \mid H_0 \text{ is true})=0.4$		$2\,\Pr(U \le 0 \mid H_0 \text{ is true})=0.1$	
	No evidence to reject H_0: $\mu_x=\mu_y$		No evidence to reject H_0: $\mu_x=\mu_y$	
	against H_0: $\mu_x \ne \mu_y$ at $^*P<0.05$		against H_0: $\mu_x \ne \mu_y$ at $^*P<0.05$	
	Barley ($m=3$, $n=3$)		Maize ($m=3$, $n=3$)	
1994	136.8 (4)	158.0 (5)	191.9 (4)	193.0 (5)
1995	112.0 (2)	133.0 (3)	138.8 (1)	160.0 (2)
1996	109.4 (1)	139.0 (5)	246.6 (6)	164.0 (3)
	$w_x=7$, $w_y=14$, $u_x=1$, $u_y=8$		$w_x=11$, $w_y=10$, $u_x=5$, $u_y=4$	
	$2\,\Pr(U \le 1 \mid H_0 \text{ is true})=0.2$		$2\,\Pr(U \le 4 \mid H_0 \text{ is true})=1.0$	
	No evidence to reject H_0: $\mu_x=\mu_y$		No evidence to reject H_0: $\mu_x=\mu_y$	
	against H_0: $\mu_x \ne \mu_y$ at $^*P<0.05$		against H_0: $\mu_x \ne \mu_y$ at $^*P<0.05$	
	Sugar beets			
1994	45.5 (4)	40.0 (1)		
1995	45.0 (3)	42.0 (2)		
1996	47.1 (5)	46.0 (6)		
	$w_x=13$, $w_y=8$, $u_x=7$, $u_y=2$			
	$2\,\Pr(U \le 2 \mid H_0 \text{ is true})=0.4$			
	No evidence to reject H_0: $\mu_x=\mu_y$			
	against H_0: $\mu_x \ne \mu_y$ at $^*P<0.05$			

x, Area estimates based on MARS-STAT: activity A; y, area estimates based on MARS-STAT: activity B (two sites: RA12/Thessaloniki and RA14/Farsala, three images per site, per year).
* Ordered observations in the pooled sample of x and y.

account the general cloudy weather conditions through the spring and summer time, a satisfactory acquisition success rate of 93% for the whole campaign has been achieved. The representativeness of the rapid estimates sample is 59 sites covered with one image (4 sites), two (12 sites), three (29 sites) or four (14 sites). Only one site (Swedish Koping; three images) is missing. Thus, given that only 12 images are definitely lost, the sampling coverage is reasonably good for Sweden, Finland, Austria, Germany and UK (all these countries are now achieving 80–90% acquisition rates) and satisfactory for the remaining countries.

In Hellas the program is operated by an organisation independent of the Ministry of Agriculture and provides annual information on the main crops cultivated in these areas. Two sites with three images per site are included in the 60 sites sampled. These sites are located in Thessaloniki (RA12: Central Macedonia) and Farsala (RA: 14: Thessalia). The area, yield and production estimates of the rapid estimates activity, obtained throughout the agricultural seasons of the previous years at a national level are published each year by Eurostat and reported in Table 6. However, the area frame sampling methodology has been applied over the whole of Hellenic territory since 1994 and therefore, any statistical comparison between the area frame sampling and the rapid estimates methodology should be based on a very small sample (one observation per year per crop). Further, the sampling techniques used in the rapid estimates method do not allow appropriately realistic assumptions upon which a parametric test should be made.

Thus, based on the information provided in Table 7, the non-parametric Wilcoxon two-sample test (Walpole and Myers, 1977) was carried out in order to see if there is a reasonable evidence for rejecting the original hypothesis of ineffective differences between the two methods of area frame sampling and rapid estimates (H_0: $\mu_x - \mu_y = 0$) against the alternative hypothesis of effective differences between these two methods (H_1: $\mu_x - \mu_y \neq 0$) at $*P < 0.05$. Note that such a test is based on the assumption that the populations from which the two samples were selected are continuous and generally non-normal. Only five cases (crops) needed to be considered, namely soft and durum wheat, barley, maize and sugar beets. The samples x and y are assumed random and consist of independent observations resulting from the application of the two methods of area frame sampling and rapid estimates respectively. Details of the main steps followed are shown in Table 7. As it appears there is no evidence to suggest the rejection of H_0 (ineffective differences) at 5% level in all the cases examined (two-tailed tests).

6. Conclusion

Agricultural statistical services are now becoming increasingly familiar with the potential of RS in the area of crop inventories at regional or local level. The introduction of RS entails the adoption of area frame sampling techniques, previously little used in Europe, where the prevailing method was surveys based on a list of farmers. This has been achieved with the endorsement of two activities of the MARS project, namely, the regional inventories and the rapid estimates, which have been reviewed here, in order to show how area frame sampling methodology was applied in combination with satellite imagery. The method developed is now fully operational in many European countries, supplementing structural surveys and compulsory censuses with an area survey generating more rapid, objective and accurate vegetation statistics. This appears to be an attractive approach, particularly for countries where there were occasional weaknesses, or where current systems did not fully meet the demands created by the new measures of the CAP. In practice, it was the countries of southern and eastern Europe that experienced such difficulties, owing partly to the large number of farmers. It was also in these areas that RS has been proven as the most effective approach and that operational programmes have been set up.

The main contribution of the work described in this paper was to provide an overview of the area frame of square segments methodology and to show how this sampling technique was implemented in the case of Hellas, in the framework of the activities A and B of the MARS-STAT sub-project. The method used sample survey results to plot crops on pre-determined geographical segments, which were then digitised using special software. Estimates of cultivated crop areas were directly extracted in accordance with statistically accepted criteria. The obtained digitised files were matched against the satellite images in order to provide comparable results with those obtained from ground surveys. The method has been fully operational since 1995. Thus, some estimates of the areas allocated to significant crops, such as soft and durum wheat, maize, cereals, sugar beets, cotton, tobacco, olives trees and vines were reported and showed an improvement in using the supervised classification method. However, new methods of classification are in progress so as to provide further improvements in the land cover estimates and thus to obtain more viable results.

Further, a revised stratification methodology was proposed and as a result, a new sample was obtained for the Hellenic regions of Macedonia and Thrace. The new sample was based on limited number of strata and it used no satellite data. However, the proposed methodology has taken into account census elements (1991), such as the portion of the agricultural land at a local community or municipality level. The statistical evaluation of the new sample was made through the development of a regression model between of the observed CV values and the percentage of land used for agricultural purposes. The model allowed the calculation of the estimated CV values as well as the precision of the area estimates. These results are compared with a previously applied model showing its relative validity and accuracy. Thus, the new methodology may be extended to cover all the Hellenic regions. Finally, the last contribution of this research was to provide a comparison of the results obtained when both the Activities A and B were applied. It appeared that there was no significant difference between the results obtained. Thus, the acreage, yield and production estimates obtained from the supervised classification methodology agree in practice with those obtained from the rapid estimates methodology.

As a final note it may be observed that the new CAP regulations related to set-side land will have major impacts on the agricultural pattern of the EU, i.e. in terms of crop (type) distribution and their relative importance in the various European regions, regarding the soil types selected by farmers for their crops and relation to crop rotation. The study of these impacts becomes a priority, because they may significantly affect the interpretations and results. In this respect the delivery and installation of the Orbital Remote Sensing of Crop Area (ORCA) system, which was developed throughout 1994 and will replace from 1996 onwards the existing satellite image processing and analysis chain of activity B is considered as a major step.

Acknowledgements

This work was supported by the Hellenic Ministry of Agriculture, Directorate of Agricultural Policy and Documentation in the framework of the MARS project.

The author wishes to express his gratitude to E. Kazazis, G. Mimidis, N. Papadopoulos and G. Levakos for their assistance during the course of this work.

References

Annoni, A., Gallego, F.J., 1992. Software for crop area estimation with remote sensing (MARS-PED) and quick digitisation of sample units (TTS). In: Office for Publications of the E.U., Luxembourg (Ed.), Conference on the Applications of Remote Sensing to Agricultural Statistics, 1991. Belgirate, Lake Maggiore, Italy, pp. 117–124.

Card, D., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogrammetr. Eng. Remote Sens. 48, 431–439.

Carfagna, E., Gallego, F.J., 1995. Yield estimates from area frame at European level. In: Office for Official Publications of the E.U., Luxembourg (Ed.), Seminar on Yield Forecasting, 24–27 October 1994, Villefranche sur Mer, pp. 237–240.

Cohran, W., 1977. Sampling techniques. Wiley, New York.

Cotter, J., Nealon, J., 1987. Area frame design for agricultural surveys. US Dept. of Agriculture, National Agricultural Statistical Surveys, Washington.

Cotter, J., Tomczac, C., 1994. An image analysis system to develop area sampling frames for agricultural surveys. Photogrammetr. Eng. Remote Sens. 60 (3), 299–306.

Czaplewski, R.L., Catts, G.P., 1992. Calibration of remotely sensed proportion of area misclassification error. Remote Sens. Environ. 9, 29–43.

Delince, J., Van de Steene, M., Terres J.M., Hawkins, A., 1993. Does accuracy of area measurements fit with the control requirements of the reform of CAP. Unpublished report of JRC, MARS project.

de Winne, P., 1994. Les Besoins de la Direction Generale VI: Agriculture. In: Office for official publications of the E.U., Luxembourg (Ed.), Report EUR 15599 EN, Conference on the MARS project: Overview and prospectives, Villa Carlotta, Belgirate, Lake Maggiore, Italy, 17–18 November 1993, pp. 17–22 (in French).

Fierens, F., Rosin, P., 1994. Filtering remote sensing data in the spatial and feature domains. In: Proceedings of the European Symposium on Satellite Remote Sensing, EC, EOS, SPIE, vol. 2315. ORA/PRO 38442, 26–30 September 1994, Rome, Italy.

Gallego, F.J., 1995. Sampling frames with square segments. In: Office for official publications of the E.U., Luxembourg (Ed.), Report EUR 16317 EN.

Gonzalez, A.F., Lopez, S., Cuevas, J.M., 1990. Comparing two methodologies for crop estimation in Spain using Landsat-tm images and ground gathered data. Remote Sens. Environ. 32, 29–36.

Hay, A.M., 1988. Global estimates from confusion matrices. Int. J. Remote Sens. 10, 1571–1573.

Hellenic Ministry of Agriculture, 1991–1994. Restructuring agricultural statistics with remote sensing. Department of Agricultural Policy and Documentation. Year Reports.

Hellenic Ministry of Agriculture, 1996. Restructuring agricultural statistics with remote sensing. Department of Agricultural Policy and Documentation. Final Report.

Hill, J., Aifandopoulou, D., 1990. Comparative analysis of Landsat-tm and Spot HRV-1 data for use in multiple sensor approaches. Remote Sens. Environ. 34, 55–70.

Meyer-Roux, J., Vossen, P., 1993. The first phase of the MARS project, 1988–1993: overview, methods and results. In: Office for Official Publications of the E.U., Luxembourg (Ed.), Report EUR 15599 EN, 1994. Conference on the MARS project: Overview and prospectives, Belgirade, Lake Maggiore, Italy, 17–18 November, 1993, pp. 33–79.

Pertigao, V., Burrill, A., 1994. The implementation of Agriculture Monitoring Systems in Central and Eastern Countries. GIS/LIS '94 Central Europe Conference, ORA 38209, 13–17 June 1994, Budapest, Hungary.

Pous, B., Reboux, D., Guerif, M., Delecolle, R., Fischer, A., De Boissezon, H., 1995. Previsions du rendement et information satellitaire: acquis de l'action IV du project MARS et nouvelles perspectives. In: Office for official publications of the E.U., Luxembourg (Ed.), Seminar on Yield Forcasting, 24–27 October 1994, Villefranche sur Mer, pp. 567–578 (in French with English abstract).

Sharman, M., de Boissezon, H., 1991. Action IV, de l'image aux statistiques: bilan operationnel apres deux annees d'estimation rapides des superficies et des rendementes potentiels au nineau Europeen. In: Office for Publications of the E.U., Luxembourg (Ed.), Conference on the Applications of Remote Sensing to Agricultural Statistics, Belgirate, Lake Maggiore, Italy, pp. 177–186, (in French with English abstract).

Terres, J., Arvain, J., 1993. Estimation statistique des erreurs de measures de surfaces agricoles pour differents supports cartographiques. Internal report, Joint Research Centre, Ispra, Italy.

Vossen, P., Gallego, F.J., Loopuyt, Ph., Perdigao, V., Dallemand, J.F., 1995. The MARS project and its extensions in central and eastern European countries. In: Office for official publications of the E.U., Luxembourg (Ed.), Seminario Esquilino, Rome, 27–29 November.

Walpole, R., Myers, R., 1977. Probability and Statistics for Engineers and Scientists, 2nd edn. Collier Macmillan International Editions, N.Y, London.