Survey Methodology, December 1994
Vol. 20, No. 2, pp. 107–115
Statistics Canada

Two-Stage Area Frame Sampling on Square Segments for Farm Surveys

F.J. GALLEGO, J. DELINCÉ and E. CARFAGNA[1]

ABSTRACT

In the MARS Project (Monitoring Agriculture with Remote Sensing) of the E.C. (European Community), area frames based on a square grid are used for area estimation through ground surveys and high resolution satellite images. These satellite images are useful, though expensive, for area estimation: their use for yield estimation is not yet operational. To fill this gap the sample elements (segments) of the area survey are used as well for sampling farms with a template of points overlaid on the segment. Most often we use a fixed number of points per segment. Farmers are asked to provide global data for the farm, and estimates are computed with a Horvitz-Thompson approach. Major problems include locating farmers and checking for misunderstanding of instructions. Good results are obtained for area and for production of the main crops. Area frames need to be complemented with list frames (multiple frames) to give reliable estimates for livestock.

KEY WORDS: Area frame; Point sampling; Segment sampling; Farm sampling.

1. INTRODUCTION

The main purpose of this paper is to present the method used to sample farms in an area frame by the MARS (Monitoring Agriculture with Remote Sensing) Project of the European Community (EC). Sampling farms is not a central activity in this project, but rather a way of bypassing the limitations of the actual capacity of satellite images, especially for yield estimation. We shall present a brief overview of the MARS Project to make up for the few existing references in statistical journals (Ambrosio 1993, Gallego 1992). Other presentations can be found in conference papers (Meyer Roux 1990, Delincé 1990, Sharman *et al.* 1992, Carfagna *et al.* 1994) or remote sensing journals (González *et al.* 1991, Gallego *et al.* 1993).

2. THE MARS PROJECT OF THE EUROPEAN COMMUNITY

The MARS Project was launched in 1988 to assess and to develop operational applications of Remote Sensing to Agricultural Statistics. It is carried out by the Institute of Remote Sensing Applications (IRSA) of the Joint Research Centre (JRC) of the EC. Most of the activities of the period 1988-1993 were divided into 4 main parts, named "actions":

(1) Regional Crop Inventories.
(2) Monitoring Vegetation.
(3) Agrometeorological Models.
(4) Rapid Estimates at the EC level.

Some work is made as well in other related fields, such as area frame sampling. We shall focus here on a sampling method used in the frame of action 1 "Regional Inventories", but we shall first say a word about the other actions.

2.1 Monitoring Vegetation

This action deals with low resolution satellite images from NOAA-AVHRR (Advanced Very High Resolution Radiometer). In these images each pixel has about 1 km^2 in the vertical of the satellite orbit. The main objectives are the development of friendly software for the pre-treatment of these images, and building a data bank with time series vegetation indexes and other indicators for about 3,000 monitoring units in the EC. These monitoring units have not yet been definitely defined. They should be geographic areas roughly between 500 km^2 and 1,000 km^2 with a more or less homogeneous vegetation or greenness index (Houston 1984, Goward 1991).

2.2 Agrometeorological Models

General and crop specific models are being currently developed on the basis of data from a network of about 650 Meteorological Observatories in Europe and surrounding areas. This model CGMS (Crop Growth Monitoring System), developed in collaboration with the WOFOST (World Food Studies Centre, in Wageningen, Netherlands), also uses other data, such as soil and elevation data, together with information on the physiology of plants (van Diepen 1989, van Lanen 1992). Remote sensing (low resolution images) will come into the picture later for the spatial interpolation of ground observed meteorological data. Parameters of the model are currently computed for each cell of a 50 km × 50 km grid.

[1] F.J. Gallego, Joint Research Centre of the European Communities, tp. 440, 21020 Ispra, Varese, Italy; J. Delincé, Commission of the E.C. DG VI, Loi 120, 4-23, 1049 Brussels, Belgium; E. Carfagna, Department of Statistics, University of Bologna, V. Belle Arti 41, 40126 Bologna, Italy.

2.3 Rapid Estimates at the E.C. Level

The main goal is giving rapid estimates of area and yield change of annual crops compared with the previous year based on a two-stage sampling scheme: 53 sites (Figure 1) of 40 km × 40 km with a sample of 16 squared segments of 700 m × 700 m (Figure 2) in each of the sites. Individual data are acquired by photo-interpretation of SPOT-XS or Landsat-TM images. An average of three images is analysed for each site with a minimum of ground information, namely a general knowledge of the dominant crops in each area. A ground survey is made for an a posteriori validation of the photo-interpretation. A monthly report (from March to November) is produced with an update of the estimates. Each report should use all the images acquired more than 15 days before.

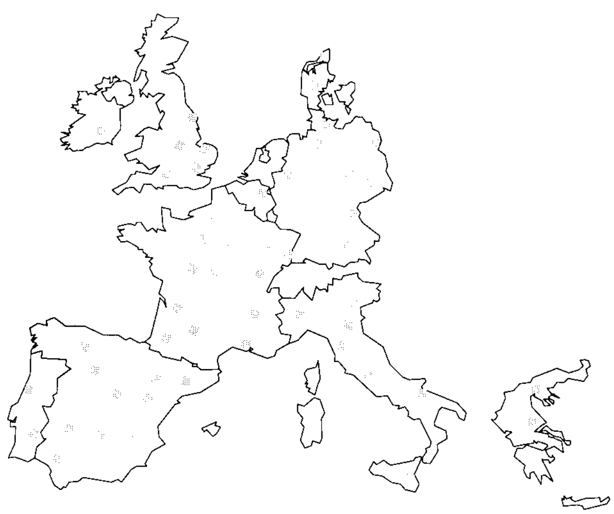

Figure 1. Sample of 53 sites for rapid crop estimates in the E.C.

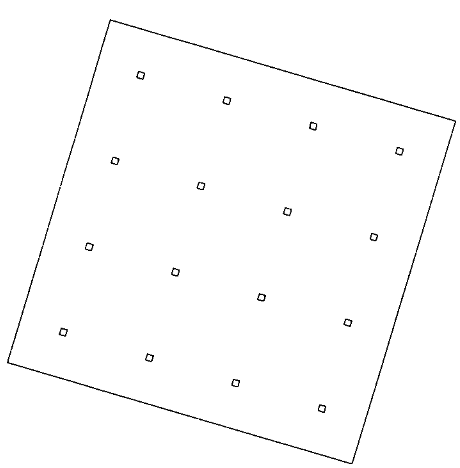

Figure 2. Segments in one site (rapid estimates in the E.C.)

2.4 Regional Crop Inventories by Segment Sampling and Remote Sensing

The objective of the action was to implement, to adapt and to assess estimation methods for crop area and production based on area frame sampling and satellite images. When this action was implemented by the IRSA in 1988 on five pilot regions of approximately 20,000 km^2 each; an absolute priority was given to annual crops: soft and durum wheat, barley, rapeseed, dried pulses, sunflower, maize, cotton, tobacco, sugar beet, potatoes, rice and soya, as well as fallow. Attention is being shared more and more by permanent crops, pastures, and non-agricultural land uses.

Since 1990 the IRSA has progressively transferred the initiative to regional or national administrations that wish to use area frame surveys based on segments. In general, the activities have been shifted to the southern countries of the EC and the former communist countries in central Europe, that have shown much interest in the method (Figure 3). In some cases, like in Italy, there is just an exchange of points of view between the national project and the IRSA.

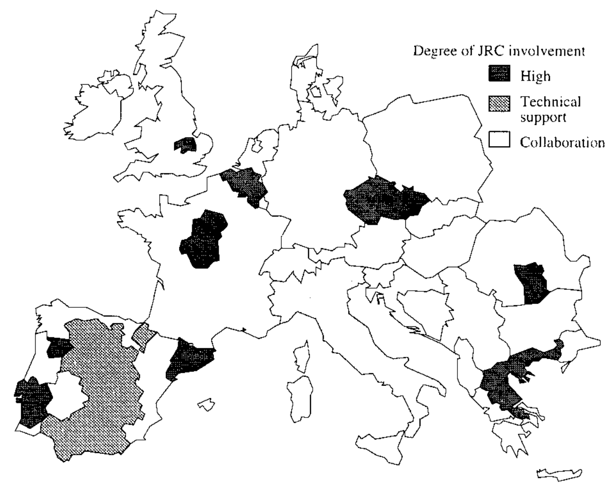

Figure 3. European regions with segment surveys in 1992.

2.4.1 Sampling Segments on a Square Grid

There are two main approaches to building an area frame based on segments: the segments can be drawn on topographic or cadastral maps following roads, rivers, or limits of fields (sometimes called cadastral segments). The sample is usually drawn with a two-stage procedure with intermediate primary sample units to reduce the burden to build the frame (Cotter 1987), which remains in any case a heavy operation.

We generally use frames based on a square grid (Gallego and Delincé 1994), which is much faster to define. Satellite images are generally (but not necessarily) used for stratification prior to sampling.

Figure 4 illustrates a small example of this kind of sample with a very simple stratification and segments of 25 ha (hectares). Sampling is systematic, repeating a pattern in square blocks. In this case the blocks have a size of 10 km × 10 km, and the pattern has 4 replicas in the most agricultural stratum (plain), 2 replicas in the hills, and one in the mountains.

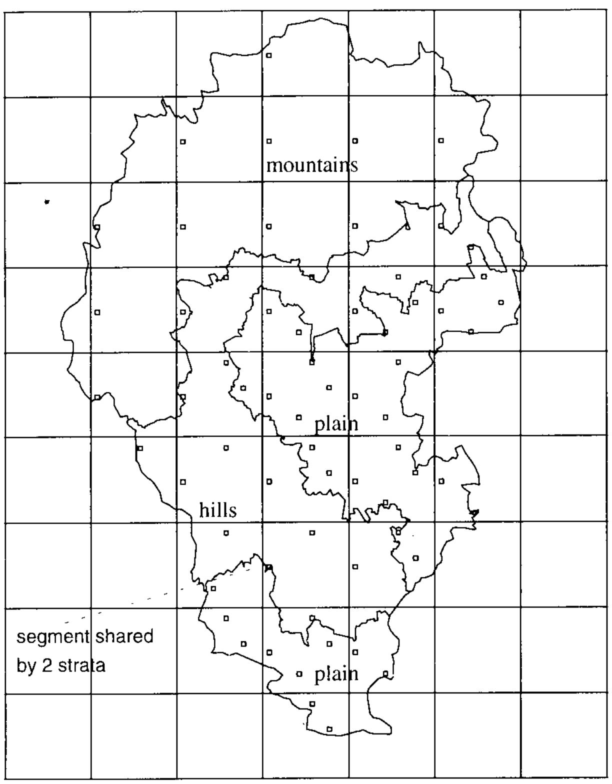

Figure 4. Example of area frame sample with squared segments and squared blocks.

The main drawback of this approach is the management of segments that fall on the boundary between two strata (Figure 5). Three alternatives are being tested for this problem: (1) adapting the stratification to the sampling grid, (2) splitting border segments into pieces that belong to different strata, and (3) keeping only the largest one among these pieces.

The most frequent non-sampling errors – shifts in location and inaccuracy in shape or size of the segment – are not strongly correlated with land use. No major influence has been found on the area estimates or their precision.

The sample pattern to be repeated in each block is drawn at random with a restriction on the distance between segments in order to avoid segments that are too close to each other. Cluster estimators can be used in this case rather than standard formulae for random sampling (Fuentes 1994, Ambrosio 1993). Systematic sampling has a risk of bias if there is a cyclic effect in the landscape with a period that coincides with the block size (10 km in the example), but this is very unlikely. The distance threshold between segments can induce an overestimation of standard errors if the spatial correlation is significantly positive for distances less than the threshold.

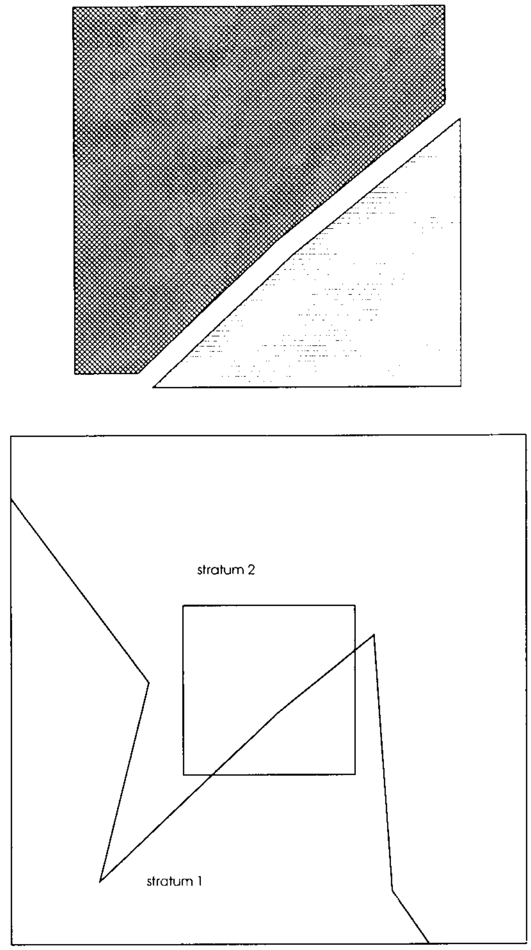

Figure 5. A segment can be split by a stratum boundary.

The size of the segments varies from region to region depending on the agricultural landscape, especially on the size of fields. In the Czech Republic, the segment size was 400 ha. For the area survey, enumerators locate the segments, draw fields on a transparent sheet placed over an aerial photograph, and write down their land use. About 5% to 10% of the segments are visited again by supervisors to check for possible errors on the ground work. Satellite images are not used either for the survey itself or for the farm survey, but they can be optionally used to improve the precision of the area estimates as described in the next section.

2.4.2 Improving Area Estimates with Satellite Images

High resolution satellite images from Landsat-TM or SPOT-XS sensors have been assessed and are still being used at moderate scale to improve the estimates obtained from the ground survey on a sample of segments. The most commonly used approach is the regression estimator on classified images. An alternative estimator based on confusion matrices has been tested with results that are very close to those of the regression estimator (Hay 1988, Gallego 1994).

The conclusions of this assessment are similar to those of the US Department of Agriculture (Allen 1990). The use of satellite images for area estimation is operational, but still too expensive for the efficiency obtained. The economic threshold can be reached by improving image processing automation, since the cost of image processing in the European market for this purpose is much higher than the cost of the images themselves. This threshold has nearly been reached with Landsat-TM images in Greece. Different conclusions on cost analysis are presented by Giovacchini (1992).

3. SAMPLING FARMS BY POINTS

For agricultural surveys in the European Community, farms are traditionally sampled from a list frame (Eurostat 1991). The list is a census of farms that exceed a certain size threshold. In many countries an agricultural census is made every 10 years and is seldom updated (if ever). Hence there may be a substantial difference between the sampling frame and the actual population at the date of the survey. The situation is worse in the central European countries of the former Eastern Block (the area between Poland and Rumania-Bulgaria), where the change of land property structure is so rapid that the census may not exist for private farms and becomes obsolete for co-operatives.

Area frames on square segments can be easily defined when the geographic borders of the region are known. A subsample of these segments is used as well for sampling farms in several countries with the help of a template of points overlaid on the segment. This has been experimentally tested in Germany, Portugal, Italy (Carfagna 1991) and Spain, and is now being regularly used in Greece, Rumania and the Czech Republic.

The template is the same for all the segments in a stratum, and usually symmetric to reduce the risk of bias due to a particular geographic location. Data are obtained only for farms corresponding to points falling on Utilized Agricultural Area (UAA).

The definition of UAA used in the field work is adapted to each national system. Farm buildings and rough pastures are included in some countries and excluded in other countries. The crucial point is that the definition used must be consistent with the definition of the column UAA used for computation (Table 1).

Table 1

Observations Generated by Points Sampled in the Segment of Figure 6

Segment	Point	UAA	Permanent Crops	Wheat		Barley	
				Area	Production	Area	Production
1	1	19	4	12	64	0	0
1	2	0	0	0	0	0	0
1	3	0	0	0	0	0	0
1	4	35	0	24	131	3	12
1	5	35	0	24	131	3	12
2	...	...	...	...	...	...	...

In the example of figure 6, point 3 fell on woodland and point 2 on a built area. They will generate two zero-valued records in the farm file. The enumerator will have to locate the farmers for the other three points. The farm corresponding to point 1 has other fields in the segment, that will be implicitly included in the survey, but the enumerator will not need to find out if these fields exist. Points 4 and 5 belong to the same farm, and it will appear twice in the farm file (Table 1).

Figure 6. Segment with a pattern of 5 points for farm sampling.

Farmers are located and asked to provide global data for the farm, including total area and production of each target crop. No question is asked about the production of each field or the set of fields inside the segment. This is not necessary because in the final formulae to compute the estimates (formulae 2 and 3 in section 4.1) the crop area or the production in the tract is not used.

The ground survey instructions are usually transferred from the JRC to National Administrations. They explain the instructions to Regional co-ordinators, who give the information to the enumerators. Instructions may be modified in some of these steps. Checking that the instructions have not been misunderstood is sometimes difficult, in part because linguistic limitations are a serious barrier to direct contact with enumerators. In some countries (*e.g.*, Spain) farmers live mainly in rather large urban nuclei and are difficult to locate; this can lead to a significant amount of missing data.

4. ESTIMATES BASED ON FARMS SAMPLED BY POINTS

We assume that the population Ω of segments is divided into strata Ω_h, $h = 1, \ldots, H$, the total population size is N segments (N_h for stratum Ω_h) and the sample size is n segments (n_h). The size of our sample of points in each segment will be K_i, previously fixed; in general we have $K_i = K$, constant across all strata, out of which F_i correspond to the farms on which these points fall. Each segment i has a total UAA surface U_i.

We have a two-staged sampling scheme. In the first stage the segment i is selected with probability $p_i = 1/N_h$ in each of the n_h trials. In the second stage the unit is not the farm but the tract (UAA in a segment, that belongs to the same farm). The tract k of segment i has an area T_{ik}. The total UAA of the farm is A_{ik} over all segments. U_i is the sum of the tracts T_{ik} in segment i.

The method presented below is closely related to the so called "weighted segment estimator" approach used in the U.S. and in Canada (Nealon 1984).

4.1 Estimates Based on Farms and Non-Farm Points

There will be $K - F_i$ observations (fictitious farms) with value 0 corresponding to points outside the UAA.

Sampling through points means that tracts are selected with replacement and with a probability p_{ik} proportional to the area T_{ik}/D_i, (the knowledge of T_{ik} is not necessary), where D_i is the size of the segment determined by the frame design. We are implicitly assuming that the surveyed region is flat. A slight bias might be introduced by the fact that annual crops are usually on more or less flat land and pastures or non-UAA are often on land with a steeper slope.

The sampling is done with replacement: a farm can be selected more than once, which gives easier formulae for variance estimation. Strictly speaking the joint selection probability that farms k and k' are in the sample $p_{ikk'} \neq p_{ikk} \times p_{ikk}$ as would be the case if the different points of the template were drawn independently, since there is usually a relatively large distance between them. We will disregard this fact in this paper.

W_{ik} will be an additive quantity for a farm, most often the production or the area of a particular crop. It is obvious that yield is not an additive variable.

Since we have no information about how W_{ik} is distributed inside the farm, we create a fictitious variable X that is uniformly distributed, and that has, by definition, the same total as W for each farm:

$$X_{ik} = \frac{T_{ik}}{A_{ik}} W_{ik}. \tag{1}$$

Estimating the totals of X and W are equivalent problems.

The two-stage version of the Horvitz-Thompson estimator for the total of X in the stratum Ω_h gives:

$$\hat{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{\hat{X}_i}{p_i} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{K_i} \sum_{k=1}^{K_i} \frac{X_{ik}}{p_{ik}} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{D_i}{K_i} \sum_{k=1}^{K_i} \frac{W_{ik}}{A_{ik}}. \tag{2}$$

This means that, even if the second stage sampling unit is the tract, we do not need to know its area nor X_{ik}, but just the global information about the farm.

The estimator is a linear function of the estimates on the selected segments. Its variance in stratum Ω_h can be estimated as (Cochran 1977, section 11.6):

$$\hat{V}(\hat{X}_h) = \frac{N_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} \frac{(\hat{X}_i - \bar{X}_h)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{K_i(K_i - 1)} \sum_{k=1}^{K_i} \left(\frac{W_{ik} D_i}{A_{ik}} - \hat{X}_i\right)^2. \tag{3}$$

The estimates for the total are:

$$\hat{X} = \sum_{h=1}^{H} \hat{X}_h \qquad \hat{V}(\hat{X}) = \sum_{h=1}^{H} \hat{V}(\hat{X}_h). \tag{4}$$

Crop areas are currently estimated from the segment survey with more objective ground data (direct observation of the enumerator on the filed), although some bias can appear due to the imperfect location of the segments on the ground. Farm surveys provide both area and production estimates, but they can have more significant bias due to non response and to a subjective tendency of the farmer that can depend on whether he is more concerned about taxes or about subsidies at the time of the survey. Comparing both area estimates, from segment survey and farm survey, can be useful to check for possible bias on the production estimate based on the farm survey.

The estimates are also possible for cattle, but the results will be presumably bad if there are a substantial number of farms without any UAA, which will not be sampled: the coverage of the area frame will not be complete in this case. On the other hand it may happen that the number of livestock does not correlate to the UAA and hence to the probability of selection. This results in inefficient estimates.

A program in C for Personal Computers has been written (Dicorato 1993) to compute estimates using this method. The main part of the program was first written to compute estimates on a segment survey.

4.2 Estimation Based Only on Farm Points

We shall mention another option that consists of using only points that fall in the UAA. In this case, we first fix F_i, the number of points that fall in UAA (often $F_i = F_h$, constant in each stratum). In segment i we observe as many points as necessary to have F_i points in the UAA. If the segment i has no UAA, one observation (fictitious farm) is added with 0 values. This is actually an implicit second-stage stratification or stratification of the first-stage units (segments) into two strata; UAA and non-UAA. The non-UAA stratum is not sampled. In this case (2) and (3) are to be adapted substituting K_i by F_i and D_i by U_i. Some inconsistency may arise in hilly areas because A_{ik} comes from the farmer's declaration and U_i from segments drawn on the ground over aerial photographs.

$$\hat{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{\hat{X}_i}{p_i} = \frac{N_h}{n_h} \sum_{h=1}^{n_h} \frac{1}{F_i} \sum_{k=1}^{K_i} U_i \frac{W_{ik}}{A_{ik}}, \quad (5)$$

$$\hat{V}(\hat{X}_h) = \frac{N_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} \frac{(\hat{X}_i - \bar{X}_h)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{F_i(F_i - 1)} \sum_{k=1}^{K_i} \left(\frac{W_{ik} U_i}{A_{ik}} - \hat{X}_i\right)^2, \quad (6)$$

the second term of (6) is null for segments with no UAA. This term cannot be computed if $F_i = 1$ because of non-response. A value 0 can be attributed, though this will lead to an underestimation of the within-segment variance, which is relatively small according to calculations made on available data (Carfagna 1992).

This approach has only been used once to resolve a misunderstanding of instructions for ground work that should have been performed following the method in section 4.1. However advantages and drawbacks of both approaches are not clear, and no systematic comparison has been made so far on the same region and same year. Using only farm points can increase the cost of the survey if the number of points per segment is to be kept constant, but the non-UAA points removed correspond to null values of W_{ik}, and their removal can result in a reduction of the variance.

4.3 Farms with Fields in Different Strata

At first sight, the estimator (2) seems to assume that a farm k that has been selected through a point in stratum Ω_h is completely included in this stratum. It is obvious that a farm can have fields in different strata, and the question arises as to whether this fact disturbs the reliability of the results.

We stress again that the variable used is not really W_{ik}, but X_{ik} defined for each individual tract. The total of W does not coincide with the total of X in each stratum, but it does in the whole region as long as

$$\sum_i T_{ik} = A_k. \quad (7)$$

Notice that A_k is identical to what we have called previously A_{ik}, where the subindex i is used only to indicate that farm k has been selected in the sample through segment i.

This identity holds on the population, regardless of the sampling procedure, if the farms are entirely inside the region and if the geometry of the ground survey document (aerial photograph) is correct.

The perturbation due to farms with fields in different regions is expected to be small because of the low proportion (generally under 1–2%) and because there is a compensation between the bias due to fields inside the region belonging to farms with the headquarters outside the region and vice versa. We are assuming that the total of W is calculated on the farms that have their headquarters inside the surveyed region.

4.4 Nonresponse

We refer here to the estimators based on farm and non-farm points (section 4.1). If a farmer does not co-operate or cannot be found, the corresponding row or rows of the input table (Table 1) are substituted with the average values of responding farms in the segment, if there are any; otherwise they are substituted with the average of responding farms for all the segments in the current stratum.

If in the second stage (sampling farms inside the segment) we consider farm and non-farm points, and give value 0 to the points that fall in non agricultural land, it is obvious that the exclusion of nonrespondents would produce a serious bias, because the zero values corresponding to non-UAA are never missing. These points are not used to compute the "average farm" values used to fill missing values. There is still a risk of bias if farmers who cannot be located or refuse to co-operate have a peculiar behaviour, *e.g.*, if they are on the average smaller or less efficient farms.

We could have considered a different way of overcoming this problem: eliminating both missing values and a proportional number of 0 values corresponding to non-UAA points. Both give the same estimate for the total, but the second solution is more uncomfortable because the sample size in the second stage is not an integer any more.

The introduction of "average farm" values will lead to a negative bias on the variance. To compensate it, the farm is not included in the sample size K_i for the computation of variances.

5. RESULTS: TWO EXAMPLES

We discuss below some results from two regions: Emilia Romagna (Italy) and the Czech Republic. In the Czech Republic, the method presented in sections 2.4, 3 and 4.1 was used; there were no missing data at all. In Emilia Romagna the general design of the survey did not follow exactly the procedure outlined above. Missing data were treated as stated in section 4.4.

5.1 Emilia Romagna 1990

In Emilia Romagna an area of 19,500 km^2 was divided into 4 strata, excluding mountainous areas. A sample of 313 "cadastral" segments (with physical boundaries) was drawn based on a two-staged procedure with primary sampling units (psu) of about 10 km^2. Segment size was approximately 50 hectares or 100 hectares, depending on the strata. 5 points per segment were drawn at random from a grid with a 50 metre step.

Out of the 1,565 points sampled: 326 were non-UAA, the farmer's address could not be found for 206 UAA points, 38 farmers were not located and 32 refused to co-operate. 963 UAA points from 285 segments had valid data, corresponding to 617 farms, some of which appear more than once in the sample.

When we think only of area estimation, the segment survey can be seen as more objective and complete, since there are no missing data and observations do not rely on farmers' answers. If we accept this principle we can have an idea of a possible bias in the farm survey by comparing with the area estimates of the segment survey. Estimates can be compared in Table 2 for the main crops in the region. Figures match well for cereals, excepting durum wheat, and permanent crops, but some problems appear for sugar beet and soya, that might be related to misunderstandings on how to declare second crops in the same year and the same field, or with a bias due to missing values. Official statistics are produced taking into account a variety of information. Durum wheat is reported separately because of the special meaning of this crop due to the significant subsidy granted by the EC to each hectare of crop.

Table 2

Results of the Segment Survey and the Farm Survey for Main Crops in Emilia Romagna (1990)

Emilia Romagna	Segment Survey		Farm Survey				ISTAT
	Area × 1,000 ha		Area × 1,000 ha		Prod. × 1,000 tm		Area
	Esti-mation	CV %	Esti-mation	CV %	Esti-mation	CV %	
Soft wheat	212	5.7	208	6.9	1,177	8	212
Durum wheat	46	14.9	48	15.2	260	14	72
Barley	43	11.2	50	17.7	184	17	38
Rice	–	–	4	59.0	23	61	6
Sugar beet	111	7.1*	96	9.6	5,474	28	119
Soybeans	76	6.0*	55	11.6	321	39	47
Vineyards	78	13.3*	76	18.7			75
Orchards	91	13.1*	96	19.7			85

* Estimate corrected by regression on classified satellite image.
ISTAT: Official statistics. No precision provided.

The coefficients of variation in the farm survey have a reasonable behaviour for cereals, but become more difficult to understand for sugar beet and soybeans. The high CV (Coefficient of variation) for the production can be due to higher yields in larger, more specialized farms.

A correction of the production estimate can be made using the difference of area estimates between the segment survey and the farm survey. A regression estimator approach might be a good solution.

Livestock is seriously underestimated (Table 3) since many livestock owners do not have agricultural land. A mixed approach was used for cattle and pigs with an exhaustive survey using a list frame for the 50 largest farms and point sampling for the rest. The procedure works for pigs, but CVs are not yet satisfactory.

Table 3

Results of the Farm Survey on Area Frame and Mixed Frame for Livestock in Emilia Romagna (1990)

× 1,000 Units	Census	Area Frame		Mixed Frame	
		Estimate	CV %	Estimate	CV %
Cattle	869	829	14	894	13
Pigs	1,876	1,312	37	1,818	27
Sheep	90	38	74		

5.2 Czech Republic 1992

Area frames seem especially useful in the former communist countries in Europe because of the rapid change of property structure. Agricultural statistics are mainly produced with no sampling error by adding the data reported by each state farm or co-operative. This procedure will collapse in the coming years. It will be extremely difficult to have an idea of the number of existing farms, and an agricultural census will be out of date before the data are elaborated. Area frames might be the best alternative.

The territory of the Czech Republic (about 80,000 km^2) has been stratified into 6 strata by photo-interpretation of Landsat-TM images. The stratification needed 15 working days for one person. In 1992, a survey was made with a sample of 417 square segments of 400 ha drawn by repetition of a fixed pattern on blocks of 40 km × 40 km. Segments were visited and area estimates obtained as explained in section 2.4.1.

Farms have been sampled using a fixed grid of 5 points in each segment. The shape of the 5-point grid was in "x" like in figure 6. This procedure gave 2,085 points: 858 non-agricultural, and the other 1,227 from 458 farms. No missing data were recorded: all the farms were identified and none refused to co-operate. This happened mainly because the old structure of large farms was still nearly intact.

Table 4 compares the results of the segment survey (direct observations on the field), the farm survey (farms sampled by points), and official statistics for the main crops in the country. Official statistics are obtained by adding figures reported by all the state farms or co-operatives. There is a moderate disagreement on area estimates for wheat, maize, and potatoes. We should not exclude a bias in farmers' answers that has to do with self-consumption of agricultural products.

Table 4

Results of the Segment Survey and the Farm Survey in the Czech Republic (1992)

× 1,000 ha	Segment Survey		Farm Survey				CSO	
	Area	CV %	Area	CV %	Prod.	CV %	Area	Prod.
Wheat	824	5.4	757	3.7	3,412	4.9	780	3,413
Barley	655	5.1	630	3.8	2,521	4.3	640	2,512
Rapeseed	140	11.6	137	6.8	310	7.5	136	296
Sugar beet	119	11.5	127	8.1	4,172	11.0	125	3,874
Maize	361	7.5	326	4.8	8,884	4.3	361	8,904
Potatoes	109	13.6	92	7.9	1,706	8.7	111	1,969

CSO: Czech Statistical Office.

The coefficients of variation (CV) of the area estimates are lower in the farm survey than in the segment survey. This is not surprising since the farm survey gives information about fields outside the segments. The 458 selected farms represent more than 15% of the total UAA in the country. The CVs for production estimates are slightly higher than for area estimates (even lower in the case of maize). This seems to indicate that the variability of yields contributes less than the variability of areas to the variability of production.

6. CONCLUSIONS AND RECOMMENDATIONS

Area frames based on square grids are a pragmatic alternative to area frames based on ground elements delimited by physical features. They are much cheaper to build and they do not seem to have major drawbacks regarding the final results. However some theoretical work is still needed to determine under which conditions the location errors due to non-physical limits have a negligible effect on the estimates.

Sampling points inside area segments provides a feasible way to build frames for farm sampling. They are extremely useful if list frames (census) are poorly updated or do not exist. Sampling a few points per segment can be much cheaper than surveying all the farms with fields in the segment. Five points per segment seems to be a reasonable choice.

Area frames alone give poor results for livestock when the number of units is not strongly correlated with Utilized Agricultural Area of the farm.

ACKNOWLEDGEMENTS

We are grateful to the various National and Regional Administrations that have collaborated for this work, and to A. Burrill and O. O'Hanlon for the kind revision of the paper. The numerous comments by the referees have been essential for the paper to become more useful to readers.

REFERENCES

ALLEN, J.D. (1990). A look at the remote sensing applications program of the national agricultural statistics service. *Journal of Official Statistics*, 6, 393-409.

AMBROSIO, L., ALONSO, R., and VILLA, A. (1993). Estimación de superficies cultivadas por muestreo de áreas y teledetección. Precisión relativa. *Estadística Española*, 35, 91-103.

CARFAGNA, E., RAGNI, P., ROSSI, L., and TERPESSI, C. (1991). Area frame: un Nuovo Istrumento per la Realizzasione delle Statistiche Agricole in Italia. *Contributi alla Statistica Spaziale*. University of Parma.

CARFAGNA, E., and DELINCÉ, J. (1992). Farm survey based on area frame sampling. The case of Emilia Romagna in 1990. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publication of the E.C. Luxembourg.

CARFAGNA, E., and GALLEGO, F.J. (1994). Extrapolating intra-cluster correlation to optimize the size of segments in an area frame. Conference on Applied Statistics to Agriculture, Kansas State University, Manhattan, KS.

COCHRAN, W. (1977). *Sampling Techniques*. New York: John Wiley and Sons.

COTTER, J., and NEALON, J. (1987). Area frame design for agricultural surveys. U.S. Department of Agriculture. National Agricultural Statistics Service.

DELINCÉ, J. (1990). Un premier bilan de l'action 1 Inventaires Régionaux du Projet Agriculture après deux années d'activité. Conference on the Application of Remote Sensing to Agricultural Statistics, Varese. Office for Publication of the E.C. Luxembourg.

DICORATO, F. (1993). AIS estimation programs. User documentation. JRC Ispra.

VAN DIEPEN, C.A., WOLF, J., VAN KEULEN, H., and RAPPOLDT, C. (1989). WOFOST: A simulation model for crop production. *Soil Use and Management*, 5, 16-24.

EUROSTAT 1991. Working party, Crop Products Statistics. Methodological reports. Document AGRI/PE/333, Luxembourg.

FUENTES, M., and GALLEGO, F.J. (1994). Stratification and cluster estimator on an area frame by squared segments with an aligned sample. Conference on Applied Statistics to Agriculture, Kansas State University, Manhattan, KS.

GALLEGO, F.J. (1992). Flächenschatzungen für einjährige Feldfrüchte mit Hilfe Fernerkundung. *Neue Wege raumbezogener Statistik. Forum der Bundesstatistik*, 20, 109-120. Wiesbaden: Statistisches Bundesamt.

GALLEGO, F.J. (1994). Using a confusion matrix for area estimation with remote sensing. *Atti Convegno AIT*, Roma, 99-102.

GALLEGO, F.J., and DELINCÉ, J. (1994). Area estimation by segment sampling. In *Euro-Courses Remote sensing applied to Agricultural Statistics*.

GALLEGO, F.J., DELINCÉ, J., and RUEDA, C. (1993). Crop area estimates through remote sensing: Stability of the regression correction. *International Journal of Remote Sensing*. 14, 18, 3433-3445.

GIOVACCHINI, A. (1992). Agricultural statistics by remote sensing in Italy: an ultimate cost analysis. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publication of the E.C. Luxembourg.

GONZÁLEZ, F., LOPEZ, S., and CUEVAS, J.M. (1991). Comparing two methodologies for crop area estimation in Spain using landsat TM images and ground gathered data. *Remote Sensing of the Environment*, 32, 29-36.

GOWARD, S.N., MARKHAM, B., DYE, D.G., DULANEY, W., and YANG, J. (1991). Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sensing of the Environment*, 35, 257-277.

HAY, A.M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9, 1395-1398.

HOUSTON, A.G., and HALL, F.G. (1984). Use of satellite data in agricultural surveys. *Communications in Statistics Theory and Methods*, 13, 23, 2857-2880.

VAN LANEN, H.A.J., VAN DIEPEN, C.A., REINDS, G.J., DE KONING, G.H.J., BULENS, J.D., and BREGT, A.K. (1992). Physical land evaluation methods and GIS to explore the crop growth potential and its effects within the European Communities. *Agricultural Systems*, 39, 307-328.

MEYER-ROUX, J. (1990). Présentation du projet pilote de télédétection appliquée aux statistiques agricoles. Conference on the Application of Remote Sensing to Agricultural Statistics. Office for Publications of the E.C. Luxembourg.

NEALON, J.P. (1984). Review of the multiple and area frame estimators. U.S. Department of Agriculture, Statistical Reporting Service, Report 80, Washington, D.C.

SHARMAN, M., and de BOISSEZON, H. (1992). Action IV: de l'image aux statistique, bilan opérationnel après deux années d'estimations rapides des superficies et des rendements potentiels au niveau Européen. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publications of the E.C. Luxembourg.