Developments in Data Analysis
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 12, Ljubljana: FDV, 1996

Modification of the Sampling Scheme for the Implementation of the MARS Project in Slovenia

Danijela Šabić, Ana Tretjak, and Irena Orešnik[1]

Abstract

The co-operation between the Remote Sensing Team at the Statistical Office of the Republic of Slovenia and the Joint Research Centre (JRC) at Ispra has been initiated in 1993 when JRC purchased the Landsat-TM satellite scanned data covering the whole territory of Slovenia. In spring 1994 the implementation of Action 1 of MARS (Monitoring Agriculture with Remote Sensing) project was approved. If was agreed that in the first year the new MARS methodology of collecting agrostatistical data will be applied on one third of Slovenia. 22 agricultural most intensive municipalities of the northeastern area were selected.

One full Landsat-TM scene and four juxtaposed quarter scenes were georeferenced with a positional error less than one pixel, i.e. 30 m. The mosaic of Slovenia with dominant reflectance values of fundamental land use classes in scale 1 : 75,000 was printed and boundaries of 6 strata were drawn and digitised. The delineating criteria for each stratum was the intensity of agricultural land.

The area of the 22 municipalities was grided into blocks of 10 km x 10 km size, and 25 x 25 segments of size 400 m x 400 m were drawn in each block. Of the 625 segments per block a number of segments defined in advance was randomly selected, thus covering a defined sampling rate per stratum. The segments were transposed on aerial photos and a ground survey was performed. Data on identified crops on fields for each segment were scanned and extrapolated to the acreage of each stratum. In addition the standard error of the estimated acreage for each crop was computed.

The obtained results will be compared with the official statistical data and the obtained differences analysed with the help of classified satellite scanned data of the same area.

In addition the segments were used for the yield survey. A grid of 10 by 10 nodes was superimposed over one segment and 4 nodes were randomly selected. This template was used on the rest of the 185 segments. The farmers cultivating the fields identified by the 4 nodes were asked to participate in the yield survey. The selected farmers have to fill-in an enquiry, giving data on the acreage of the land use of their farm and data on yield of selected crops. The survey will be concluded at the end of October. The obtained data will be extrapolated to the area of each stratum.

[1] Statistical Office of the Republic of Slovenia, Vožarski pot 12, 1000 Ljubljana, Slovenia

1. Introduction

In Slovenia the data for crop inventory and yield are now collected by the Statistical Office of the Republic of Slovenia. The data for the public sector are collected with the census while for the private sector by estimating the statistical estimation units. The accuracy of the obtained results has never been calculated and it is suspected that a strong subjective bias is present in the obtained results.

In co-operation with the Joint Research Centre (JRC) from Ispra in Italy it has been agreed in December 1993 to start with the implementation of the MARS project (Monitoring Agriculture with Remote Sensing). The sampling scheme of the MARS project is based on the stratification of the area of interest, therefore an uniformly update land use map of Slovenia in scale 1 : 75,000 of Slovenia was needed.

Landsat-TM data covering the whole territory of Slovenia, acquired by JRC-Ispra

Orbit path/row	Type of scene	Scanning date in 1993
191/028	Q-floating*	04 July
191/028	Q-floating*	04 July
190/028	Full scene	26 May
189/028	Q-1	20 June
189/028	Q-3	20 June

Q-floating* is a quarter scene of the full scene shifted E/W from the nominal image centre

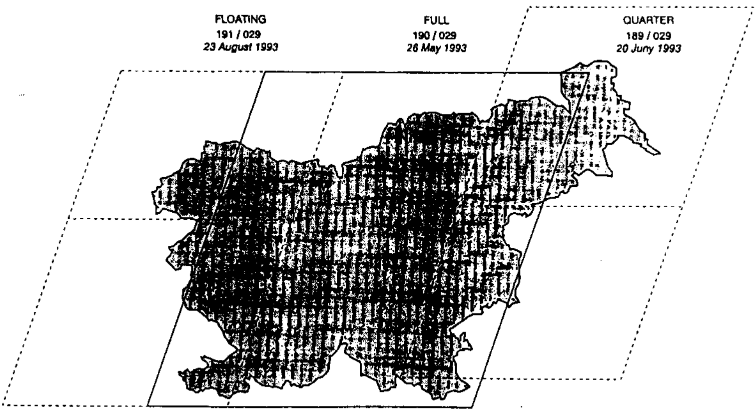


Figure 1: Landsat-TM scenes covering Slovenia

2. Compilation of the land use map

The Landsat satellite scanned data were the only available source of data that would comply to the above mentioned needs and are in addition purchased in a digital mode. A standard Landsat-TM satellite scene, i.e. full scene, covers an area of 185 km by 180 km with a ground resolution of 30 m. From that ground area unit, called the pixel, spectral measurements of reflected light are recorded on board of the satellite in 7 wavelengths called bands or channels.

In autumn 1993 the Landsat TM data covering the whole Slovenia were obtained from JRC. Different acquisition dates were selected due to cloud coverage of the territory of interest determined on the Quick-Look prints for the period from 1 May to 1 November 1993 (see Figure 1).

Table 1

Number of GCP and RMS errors for all 5 Landsat-TM scenes

Scene	No GCP	X-RMS	Y-RMS	Total RMS
191/028 Q-F 1	45	0,64	0,75	0,98
191/028 Q-F 2	49	0,70	0,68	0,98
190/028 Full	144	0,71	0,70	0,99
189/028 Q-1	121	0,67	0,73	0,99
189/028 Q-3	59	0,77	0,60	0,98

GCP is Ground Control Point
RMS is Root Mean Square error; RMS = $(X_r-X_i)^2+(Y_r-Y_i)^2$;
X_i, Y_i = input co-ordinates
X_r, Y_r = transformed co-ordinates

Location accuracy for selected points of the mosaic

Pixel co-ordinate	mosaic	Gauss-Kruger co-ordinates map
950/2618	54 03 664/51 14 762	54 03 675/51 14 775
1335/3748	54 15 244/50 80 862	54 15 275/50 80 900
3798/2169	54 89 104/51 28 232	54 89 125/51 28 175
3777/2192	54 88 474/51 27 542	54 88 450/51 27 525
3808/4608	54 89 404/50 55 062	54 89 450/54 55 050
4876/1098	55 21 444/51 60 362	55 21 475/51 60 350
5531/3607	55 41 050/50 85 092	55 41 094/50 85 100
6496/1273	55 70 044/51 55 112	55 70 075/51 55 100

Average difference in m: x = 30 m ; y = 21 m

Each scene was georeferenced to the Gauss-Kruger projection with ground control points (GCP) identified on the TM data and on maps in scale 1 : 25,000. The overall root mean square (RMS) error for each scene was less than one pixel (see Table 1). The georeferenced scenes were mosaiced. Co-ordinates of a set of 8 points in valleys, on plains and on mountainous regions were taken and compared with map co-ordinates (Tretjak and Šabić, 1993). The obtained average location discrepancy was less than one pixel (30 m), i.e. less than 1/2 mm on a map of scale 1 : 75,000.

To extract the territory of Slovenia from the TM mosaic the database of administrative boundaries was used and the file with the official state boundary applied as a cutter (see Figure 2). As the state boundary has been digitised from maps 1 : 5,000, the consistence of the boundary with the georeferenced satellite scanned data was visually inspected; on tracts where the boundary passes along rivers, a perfect fit has been confirmed.

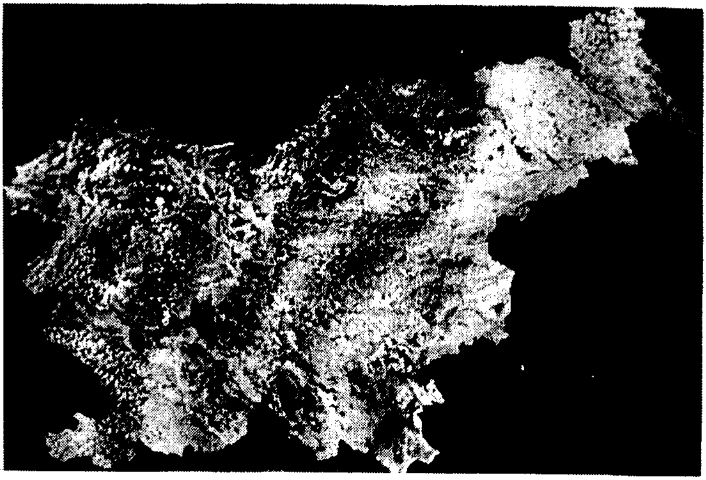


Figure 2: Final output of the mosaiced Landsat-TM/93 data covering the territory of Slovenia

3. Stratification

The combination of the reflected wavelengths from the three infrared bands (band 4 : 0.76-0.90 un; band 5 : 1.55-1.75 un; band 7 : 2.08-2.35 un), showed the best contrast between agricultural areas and other land use features and was therefore chosen to be used for the map output of the mosaic in scale 1 : 75,000 (Bellow, 1991; Harris, 1990). Strata boundaries ware drawn on the overlaid transparencies, digitized and acreage of each strata calculated.

The first criteria to define the agricultural strata was the different intensity of reflected values, resulting in bright colours on the plotted output. The frequency and size of these bright agricultural areas determined the intensity of agricultural activities of that region and were the second criteria to distinguish among agricultural intensive, medium intensive, less intensive and eventually non-agriculture areas, i.e. strata.

The 30 m sized pixel of Landsat-TM satellite also enabled to distinguish large-size fields (>5 ha) from small-size fields within the agricultural stratum.

Finally the whole territory of Slovenia was divided into 6 strata :

1. Intensive agriculture with small-size fields wheat, maize, sugar beer, meadows.
2. Intensive agriculture with large-size fields: wheat, maize, sugar beet, meadows.
3. Intensive agriculture with hops: predominantly hops and to some extend wheat, maize, meadows.
4. Medium intensive agriculture: predominantly vineyards, orchards, meadows, pastures.
5. Less intensive agriculture: mostly meadows, pastures.
6. Non-agriculture: mountainous with forest, heathlands, rocks.

As the size of the fields in Slovenia is very small and as far as crops are concerned the fields are very heterogeneous, it was decided to implement the MARS project in the first year only to one third of Slovenia. Our work was concentrated to the northeastern part of Slovenia, being the most agriculturally intensive region. In delineating the boundary of the territory of interest for MARS-94 project two rules were observed:

- not to cross the municipality's boundaries in order to retain the possibility to compare the results with the official statistics;
- not to cross the agricultural regions which were defined in 1993 with the inspection of aerial photos by the Remote Sensing Team.

Finally 22 municipalities from the northeastern part of Slovenia were selected, covering an area of 6,683 sq km (see Figure 3).

Figure 3: 22 minicipalities of agricultural most intensive region of Slovenia selected for the implementation of the MARS-94 project

4. The sampling scheme

The sampling scheme most frequently applied in the MARS projects is the so called 'systematic aligned sample with threshold distance'.

In the first step the size of the sampling unit, i.e. the segment, is determined as to comprise in the average 20 to 30 fields (Gallego and Delince, 1992). In the European countries the segment size ranged from 500 m x 500 m up to 2 km x 2 km. With the inspection of aerial photos covering the area of interest we decided that according to the average field size a segment size of 400 m x 400 m would be appropriate for Slovenia.

The performance of the MARS project is followed by superimposing a grid of squared blocks on the stratification map. Blocks are designed in the way that between 100 and 400 segments are included within a block; whenever possible the different sampling rates in different strata should be covered with an integer number of segments (Avernier et al., 1992). In addition no agricultural phenomenon should have the same pattern of periodicity as the bloc length. For Slovenia blocks of size 10 km x 10 km with 25 by 25 segments were designed.

In the so called "presampling stage" (Avernier et al., 1992) in one block 6 segments covering approximately 1% of the area were randomly selected and exactly the same pattern of selected samples repeated in all blocks (Figure 4 and 5).

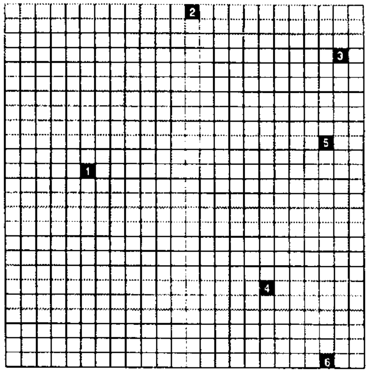


Figure 4: Template of selected segments in a block

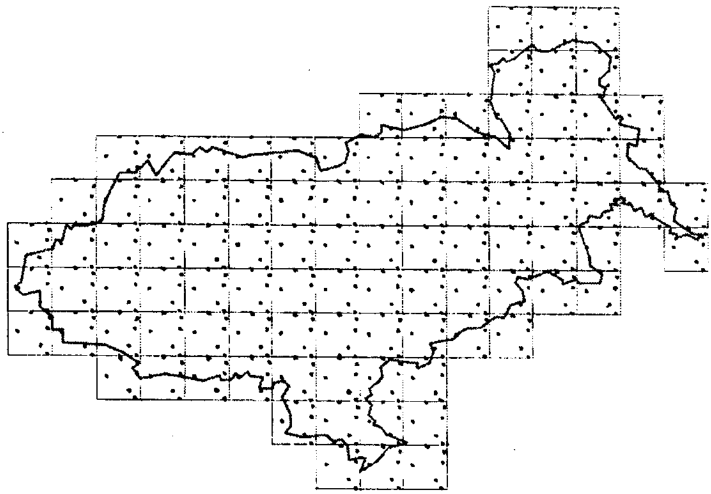


Figure 5: The distribution of segments in the 'presampling stage'

Depending on the percentage of agricultural land in each stratum, different sampling rates were selected. The sampling rates were defined following the prescribed MARS methodology (Gallego and Delince, 1993) - i.e.: 1% for the intensive agriculture stratum, 0.5% for medium intensive agriculture stratum, 0.3% for less intensive agriculture stratum and 0.2% for non-agriculture stratum - resulted in different number of segments per stratum. Only in blocks of the intensive agricultural stratum all the 6 segments were kept in the scheme; in the medium intensive agriculture stratum the first three selected segments remained in each block; in the less intensive agriculture stratum only the first two selected segments remained and in the non-agriculture stratum only the first selected segment remained in the final sampling scheme (see Figure 6).

Figure 6: Final allocation of segments on the segments on the part of Slovenia on which the MARS-94 project was implemented

A special problem are the segments lying on the boundaries between two strata. There are three approaches to solve this problem:

- the stratum boundary is modified so that the segment lies completely in one stratum;
- the segment is shifted to fall within one stratum;
- the segment is assigned to the stratum where more than 50% of segment area lies.

In the last case only the part of the segment that is lying in the assigned stratum is taken in consideration; these data are extrapolated to the area of the whole segment. On the basis of the experience gained by MARS projects performed in other European countries, it was decided to apply the third solution for MARS in Slovenia (Gallego, 1994).

Finally 185 segments were obtained to cover the area of Slovenia included in the MARS 94 project:

- 58 segments in stratum I: intensive agriculture with small-size fields;
- 9 in stratum II: intensive agriculture with large-size fields;
- 9 in stratum III: intensive agriculture with hops;
- 57 in stratum IV: medium intensive agriculture;
- 31 in stratum V: less intensive;
- 21 in stratum VI: non-agriculture.

The two strata with only 9 segments (stratum II and III) were closely examined. As stratum II does not consist of one homogeneous polygon but of few very longish, narrow polygons, it was obvious that most of the segments will not lie entirely inside the stratum. The main difference between the first and the second stratum was only in the size of the fields and not in different crops that occur. Therefore it was decided to merge strata I and II (see Figure 7 and Table 2).

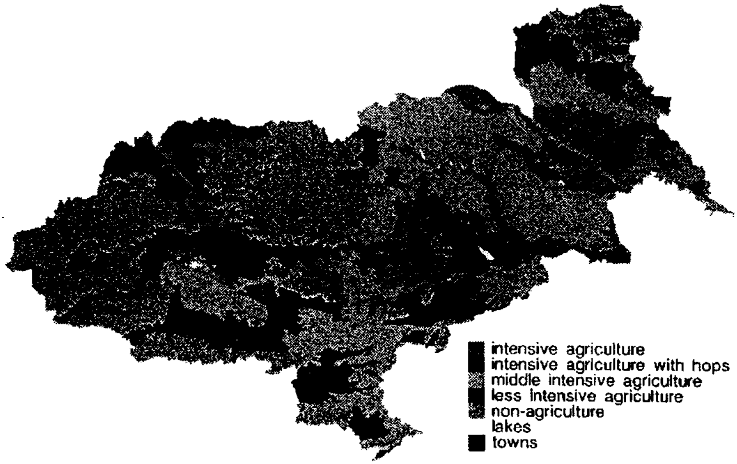


Figure 7: Final stratification of the part of Slovenia on which the MARS-94 project was implemented

Stratum III, which also had only 9 segments, remained unchanged as it consists of one large polygon with hops fields being the predominant crop of this region.

In the stratum VI (non-agricultural) the 21 segments lying on hilly and not easily accessible regions were delineated on aerial photos and photointerpreted.

5. The acreage ground survey

On the rest of the 164 segments (Table 2) the acreage ground survey was performed in the field. The inventory of segments was done in the period from 23 May to 6 June

1994. Maps were used as the basis for segment location and transparencies overlaid over enlarged aerial photos (1 : 5,000) were used for recording the inventory data.

The data on the transparencies were coded using the CRONOS nomenclature (Figure 8). Visual inspection was performed on each of the 164 segment transparencies. 10% of randomly selected segments were additionally visited on the field by a control group of enumerators and a second inventory was performed. Minor differences were obtained as far as crop identification and field boundaries are concerned. The results of the first inventory of all segments were accepted for further analysis.

Table 2

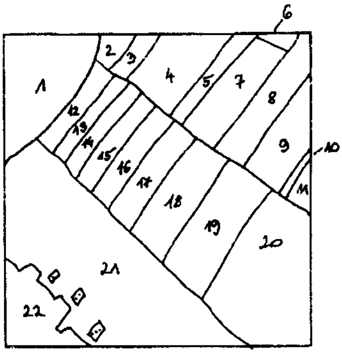
The acreage of 7 strata n MARS-94 for Slovenia

N Stratum	Acreage ha	%
1. Intensive agriculture	122 352,39	18,3
2. Intensive agri. with hops	14 770,82	2,2
3. Medium intensive agri.	186 840,90	27,9
4. Less intensive agri.	146 857,40	22,0
5. Non-agriculture	139 759,00	29,0
6. Lakes	1 212,57	0,2
7. Towns	2 502,64	0,4
Total (22 municipalities)	668 295,72	100,0

The segments' allocation per strata

N Stratum	No of selected segments	Sampling rate
1. Intensive agriculture	67	0,87
2. Intensive agri. with hops	9	0,97
3. Medium intensive agri.	57	0,49
4. Less intensive agri.	31	0,33
5. Non-agriculture	21*	0,29
Total	185	

* = an inventory of these 21 segments was made by photointerpretation



No	Code	Description	No	Code	Description
1	0061	Broad leaved forest	12	1123	Soft winter wheat
2	1200	Grain maize	13	1123	Soft winter wheat
3	1123	Soft winter wheat	14	1360	Potatoes
4	1370	Sugar beet	15	1200	Grain maize
5	1360	Potatoes	16	1163	Winter barley
6	1123	Soft winter wheat	17	1370	Sugar beet
7	1163	Winter barley	18	1123	Soft winter wheat
8	1200	Grain maize	19	1200	Grain maize
9	1123	Soft winter wheat	20	1370	Sugar beet
10	2721	Permanent grassland	21	2721	Permanent grassland
11	1163	Winter barley	22	0080	Urban

.	0004	Kitchen garden
..	0004	Kitchen garden
...	0004	Kitchen garden

Figure 8: Drawn and coded transparencies with CRONOS nomenclature

The transparencies were scanned with an 100 dbi/inch scanner. The scanner has a scanning error caused by line thickness up to 1.5 ha on segments with 40 fields. That will have to be improved with some software modifications.

The obtained acreage for all fields in a segment was the input for the program AREAESTP/JRC. The output of that program is the acreage of crops of interest with the standard error and the coefficient of variation; in addition, for the whole region the estimator of the efficiency of the stratification is produced. The estimator of the efficiency of the stratification - that is the ratio between the estimate of the variance that would have been obtained without stratification and the estimate of the actual variance - ranges from 0.83 to 2.41 for permanent grassland and soft wheat respectively, with values for other crops somewhere in-between.

Acreage estimations of MARS-94 and the statistical data

		MARS - 1994			Official statistics	
Crop	Code	Surface in ha	Standard error	Variant. coeff.	Surface 1994[1)]	in ha 1993
Soft wheat	1120	33105,33	3520,47	10.634	28505	28993
Barley	1160	10368,00	1339,62	12,921	7163	5493
G. maize[2)]	1200	49432,24	3884,57	7,858	58265	59891
Potato	1360	5679,15	768,60	13.534	8498	11900
Sugar beet	1370	5797,60	2045,50	35.282	4901	3499
Fodder beet	1381	1816,91	290,56	15.992	2693	2126
Pumpkin	1480	1460,77	316,15	21.643	2043	1293
Orchards	2040	8019,58	1545,23	19.268	17524	22144
Vineyards	2410	7515,00	1707,38	22.720	10144	9763
Perm. grass	0002	159324,71	11906,52	7.473	---[3)]	155823

1) Preliminary results
2) Includes green maize
3) Due to denationalisation not all data reported. i.e. missing data

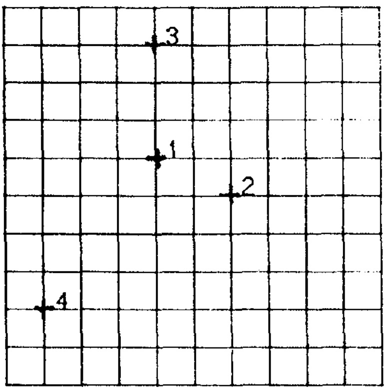


Figure 9: Template with selected nodes in a segment

The obtained results for the main crops were compared with the official statistical data of the year, which were at that time still the preliminary results (see Figure 9). The obtained differences are discussed in more details in the conclusions of this paper.

6. The yield survey

The same segments as selected for acreage survey were used for the yield survey.

A 10 x 10 nodes grid was superimposed over one segment. From the 100 points/nodes four have been randomly selected (Figure 9). The farmers cultivating fields identified by the four nodes were asked to participate in the yield survey. Up to 25 October the identified farmers will be interviewed on yield of crops that are considered as main crops of Slovenia: wheat, rye, barley, oats, grain maize, green maize, perennial green fodder, permanent grassland, potatoes, sugar beer and hops.

In addition, data on acreage of land use categories will be surveyed on these selected farms. This kind of survey is specially suitable where no exhaustive list of farmers is available (Gallego and Delince, 1993).

During the implementation of the above described procedure three special cases can occur:

The land selected by the point might not be agricultural land. In such a case no farm would be chosen.

The identified farmer might refuse to co-operate and so he contributes to the non-response error. At the time of designing the sampling scheme for the yield survey we did not have any data on the expected non-response rate. To be on the safe side an additional set of randomly selected points was prepared in order to be used for replacement. The replacement would be performed under the following rule: the first selected points have been numbered from 1 to 4, the replacement points with numbers 5 to 8. Point number 1 could be replaced only with point number 5, point number 2 with point number 6, etc..

As the identification of farmers already started during the ground acreage survey, some data on the possible non - response rate have been obtained up till now.

From the not yet completed list of farmers who have been identified and approached a 5% non-response rate has been estimated, which is according to literature considered to be low (Deeming, 1950).

The land under two ore more selected points belongs to the same farmer. In such a case the farmer will be interviewed only once, but his data will be included in computations as many times as many points he represents.

As already mentioned, most of the farmers have already been approached during the acreage ground survey and data on farmed acreage as well as sown area of the main field crops collected. After 25 October the data on the inquiry forms will be manually verified, transferred to the computer and data editing performed by the Blasé system, which enables computer assisted data collection and processing (Kozjek, 1994).

The obtained results will be extrapolated on the area of Slovenia and included in the MARS 94 project by processing the data with the YSPROSUR/JRC program.

7. Conclusions

The differences between MARS results and official statistics are difficult to explain because the accuracy of the official statistics has never been calculated. In order to be able to analyse the obtained differences, an additional analysis of Landsat-TM/93 data is now in progress. The ground truth data as well as the segments' inventories made during the field work in 1993 - both collected for the classification of satellite scanned data - will be in addition used in the analysis which will simulate the MARS project procedure as if it had been performed in 1993. The simulated MARS-93 results and the official statistical results of 1993 will be compared with Landsat-TM/93 classified results. The later considered as reference values for 1993.

When initiating the MARS project, we presumed that the small field size, the variety of different crops and the biased estimating practice of the official estimators used up till now, might result in large differences between the obtained results. As both methods - i.e. MARS 94 and the official estimating - are based on estimating, it was impossible to explain which of the applied methods results in better overall estimates of the real conditions in the field. Therefore it was decided to photointerpret part of stratum with hops in order to be able to perform an accurate study of possible sampling schemes. The photointerpreted data represent an inventory of that area and can be treated as ground truth data, i.e. reference data used for an objective verification of results obtained by different estimating methods and different sampling schemes.

On the area covered by photointerpretation the size of segments has been changed from 400 m x 400 m to 300 m x 300 m and the number of segments from covering 0.8%, 1.6%, 0.94%, 1.53%, 1.88% to 3.7% of the analysed area.

The preliminary results of this study indicate that for the next year the size of the segment will have to be diminished and the number of segments augmented. Good result having a coefficient of variation between 1-2% could most probably be obtained using a 300 m x 300 m size segment with at least a 3.7% ground coverage. Unfortunately, the application of this sampling scheme would at least triple the costs of the performance and most probably a less accurate but cost effective result will have to be accepted.

References

[1] Avernier D., Perdigao V., and Terres J.M. (1992): *Inventaire regionaux agricoles.* Institut des applications de la teledetection, Centre commun de recherhe, Ispra, Italy: CECA-CEE-CEEA, Bruxelles, Luxembourg.

[2] Bellow M. (1991): *Comparison of sensors for corn and soybean planted estimation.* USDA, Research and Application Division, 1-23.

[3] Deming W.E. (1950): *Some Theory of Sampling*. Dover publications, Inc., New York, 33 - 36.

[4] Gallego F.J. and Delince J. (1993): *Area Estimation by Segment Sampling*. MARS project; Institute for Remote Sensing applications, JRC, Ispra, Italy.

[5] Gallego F.J. (1994): *Personal Communication*. JRC, Ispra, Italy.

[6] Harris J.M. (1990): *The comparison of emulated multispectral scanner data sets*. USDA, Research and Application Division, 1-11.

[7] Kozjek P. (1994): *First experience with the Blaise system and CAI at the Statistical Office of the Republic of Slovenia*. Paper presented at the International Conference on "Satistical Data Analysis and Data Collection", 19-21 September 1994, Bled, Slovenia.

[8] Tretjak A. and Šabič D. (1993): Compilation of an agricultural GIS for the needs of agrostatistics; case study Kranj - The Republic of Slovenia. Paper presented at the International Symposium on "Operationalisation of Remote Sensing", 19-23 April 1993, ITC, Enschede, The Netherlands.

Abbreviations

AREAESTP/JRC: Agricultural Area Estimation program; developed by the Agricultural Information Systems group at Joint Research Centre (JRC) at Ispra, Italy.

CRONOS: EU nomenclature for the Ground Survey.

GCP: Ground Control Points.

JRC: Joint Research Centre at Ispra.

MARS: Monitoring Agriculture with Remote Sensing.

RMS: Root Mean Square error.

TM: Thematic Mapper; scanner on board of Landsat satellite.

YSPROSUR/JRC: Yield Surface Production Survey program; developed by the Agricultural Information Systems group at Joint Research Centre (JRC) at Ispra, Italy.