New Approaches in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 16, Ljubljana: FDV, 2000

Quality Assessment of the Statistical Land Cover/Use GIS of Slovenia-State'97

D. Šabić, E. H. Lojović, and A. Tretjak[1]

Abstract

Accurate land cover and land use data on regional level are required for agricultural statistics, national agricultural policy and monitoring environmental changes. In 1995, the first choropleth digital land cover map of Slovenia has been produced from Landsat-TM/93 satellite scanned data by stratification with a minimum mapping unit of 20 hectares of uniform land cover. The obtained digital map confirmed the long suspected but never estimated high increase of the areas under forest and built-up areas on the account of agricultural land. The improvement of the choropleth digital land cover map resulted in a statistical Land Cover GIS of Slovenia-state'93 that revealed the incoherence of the used auxiliary georeferenced databases. The follow up of the work is oriented to an updated statistical Land Cover/ Use GIS of Slovenia-state'97 on a higher level of accuracy using a new set of Landsat-TM/97 data with a 30m x 30m resolution and Spot-Pan/96-97 orthorectified satellite scanned data with a 10 m x 10 m resolution. The quality of the compiled statistical Land Cover/Use GIS of Slovenia-state '97 will be estimated using systematically distributed ground truth segments of size 3,000m x 3,100m. Areal and point methods of quality verification will be analysed regarding the precision of the results, time consumed and costs. The method chosen for the quality estimation of the produced GIS will result in the assignment of the omission and commission error in respect to the land cover category or land use class to each misclassified pixel.

1 Introduction

A Geographic Information System (GIS) is a multidimensional database system designed to input, store, retrieve, manipulate and analyse layers of georeferenced and thus scale dependant data, to produce interpretable information. The end product of a GIS is a numerical map, that comprises information on the location of

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

the phenomena under study, their mutual relation in space and in time and their occurrence in measured unites. The GIS product can be presented as a hardcopy map with tabulated statistics or as a computer file that can be used as a new input layer in the compilation of another specific GIS (Berry, 1987; Marble, 1990; Griffith, 1996). Each data layer used in the elaboration of a GIS should comprise information on the positional and thematic precision and accuracy of its elements in order to be able to analyse and control the space related error propagation during the merging process of layers into a GIS.

For the needs of agriculture statistics, a uniformly produced digital land cover map of Slovenia has been compiled in 1995 from Landsat-TM/93 satellite scanned data by stratification process with the minimal mapping unit of 20 hectares. This first choropleth digital land cover map of Slovenia confirmed the long suspected but never estimated high increase of the areas under forest and built-up areas on the account of agricultural land (Šabić et al., 1995).

In 1996 it was decided to improve the content of the existing land cover map of Slovenia. First, the nomenclature was elaborated. No satisfactory nomenclature that would clearly separate land cover categories from land use classes using only Landsat-TM data has been defined (Land use statistics 1996), since the Landsat-TM satellite scanned data give information only on reflected values of the land cover and no information on whatever use of the feature whose reflectance value is captured within the 30m x 30m pixel(s). Next, it was decided to use other official georeferenced databases that would help to determine the following land cover categories and land use classes:

- wooded land cover (parks and recreation places within urban areas excluded),
- agriculture land use, (parks and recreation places within urban areas excluded),
- water land cover,
- bare rocks as land cover
- built-up as land cover; within that land use of:
- areas under houses with yards
- areas under roads
- areas under railways and railway stations
- areas of recreation under vegetation
- masks of larger built up places
- other: in the legend of the map assigned 'undefined' since it will comprise land cover or land use such as dumping grounds, gravel pits, quarries, etc.

The descriptions of defined land cover categories and land use classes are given in the published publication: Rapid Reports No. 42/98.

2 Used georeferenced layers

As the base map the georeferenced mosaic of Slovenia has been used. It has been compiled from Landsat-TM/93 satellite scanned data with the positional root mean square (RMS) error less than a pixel, i.e. the distance between the real georeferenced location of a point and it's location on the mosaic. These data were the only reliable source of the state of land cover in 1993 uniformly covering the whole area of Slovenia. The georeferenced mosaic was produced on the Statistical Office of RS.

From other Governamental offices the following official georeferenced data for state 1993 were obtained:

- digitised boundaries of forest: derived by photointerpretation from airphotos in scale 1:5,000 or 1:10,000 by the Institute of Forestry of RS in the period from 1983 to 1989. Only 14 maps of 202 maps were declared as being in the final form as far as the thematic content of the delineated polygons was concerned;
- draft version of digitised boundaries of water, derived from maps in scale 1:5000 by the Hydrometeorology Office of RS;
- DMR-100, digital terrain model with a spatial resolution 100m by 100, derived from maps in scale 1: 25,000 by the Geodetic Survey of RS, with the declared altitude precision within 1m (not valid for the height over forests);
- centroids of houses (i.e. centre of gravity of each building) assigned to each building in the field by an official geodetic surveyor. Co-ordinates of the centroids were determined on maps in scale 1: 5,000 with a stated 1m precision and the determined co-ordinate manually submitted into the existing database that has been initiated in 1977, is operational from 1980 and is updated on a daily basis;
- vectors of roads were obtained officially from the Directorate of Roads RS, which digitised roads from maps in scale 1: 5,000 or equivalent materials. Attribute table with the description of the type of road accompanied the vectored data;
- vectors of railways were obtained from the Slovenian Railways. The data were derived from maps in scale 1: 25,000;
- register of digital administrative boundaries digitised in 1980 by the Geodetic Survey of RS from maps in scale 1: 5,000 with a stated 1 m precision and is updated whenever administrative changes are officially reported.

None of the used georeferenced data had any other information on their quality.

3 Compilation of the land cover/use GIS

The primary objective of the compiled Land Cover/Use GIS-state'93 was to obtain the information on the location and area of the main land cover categories and on agriculture land use classes: plantations, grassland, arable land and transitional class of agricultural land being in the process of grown by forest. In order to achieve that goal all the above mentioned layers of georeferenced data were prepared for the merging process (Rapid Reports no. 42/98). Merging the georeferenced layers turned into a tedious work of cleaning the illogical intersections and correcting the boundaries of the polygons where the declared land cover category did not correspond to the land cover category identified unequivocally by satellite scanned data (Ivačič et al., 1997). The work had to be thus limited only to the land cover categorisation. The obtained results were reported on the level of 12 statistical regions (similar to NUTS-3).

The part of the compiled Land Cover/Use GIS-'93 covering 21 cadastral municipalities i.e. 34km x 26km, was verified by comparing it with photointerpreted results of the same area obtained from airphotos in scale 1: 17,500 from the same year. The photointerpreted results were considered as ground truth data. The results of this analysis confirmed the structural distributions of hectares per land cover categories of the Land Cover GIS/Use-state'93:

- 57% of Slovenia is forest,
- 38% of Slovenia is agricultural land,
- 5% of Slovenia is built-up areas.

The main experience gained from the compilation of the Land Cover/Use GIS of Slovenia-state '93 was that the existing georeferenced data bases used in the compilation of the GIS did not meet the declared standards of the scale they were produced from. Merging these data did not result in a new data layer that was expected to be used for an improved classification of Landsat-TM data into land cover categories and land use classes.

4 Updating the land cover/use GIS

The objective of the follow up project is to update the existing Land Cover/Use GIS of Slovenia from the state in 1993 to the state in 1997 taking into account the experiences gained compiling the Land Cover/Use GIS of Slovenia-state'93 (Lojović et al., 1998).

In addition to the updated georeferenced databases used in the compilation of the Land Cover/Use GIS of Slovenia-state'93 the following sets of raster data are used:

- set of Landsat-TM/97 data covering the whole territory of Slovenia,
- orthorectified Spot-Pan/1996-97 scenes covering the whole territory of Slovenia. The Spot-Pan data were rectified at Spot-Image with a precision that will enable Spot-Image the compilation of a DTM-20m with a decimetre precision.

The Landsat-TM/97 data were georeferenced to the Gauss-Krüger projection with the RMS error of less than half of a pixel. Areas covered by the Spot-Pan rectified images have been cut from Landsat-TM/97 and Landsat-TM/93 mosaic and resampled to pixel size of 10m x 10m and additionally georeferenced to the Spot-Pan/97 scene with RMS error less than 10m. Following the procedure established during the elaboration of the Land Cover GIS of Slovenia-state'93 the updated Land Cover/Use GIS of Slovenia-state'97 is now in the compilation process.

5 Quality estimate of the compiled land cover/use GIS

Our attempt is to classify the new data into land cover categories and land use classes with a known locational and thematic quality.

A general visual verification of the thematic content of the polygons will be performed using the airphotoes in scale 17,500 from the latest Cyclical Aerial Survey.

The accuracy of boundary locations and the thematic content of the polygons of the categories or classes will be estimated by comparing the obtained results with interpreted segments of size 3,000m x 3,100m on maps in scale 1: 25,000. The content of these segments will be updated from the latest airphotos in scale 1:17 500 or by field work and therefore considered as reference values (Leung. 1998; Wong, 1996).

The interpreted segments will be placed into the centres of all those maps 1:25,000 that have more than 75% coverage of Slovenia i.e. 156 segments or 7% of Slovenia will be eventually interpreted for the estimation of the positional and thematic quality of the end product.

Two methods have been considered to be used for this process: the analysis of polygons and the analysis of points. When using polygons, intersections of the same land cover categories or land use classes obtained from the GIS and from the photointerpretation present correctly classified areas. The obtained results can be presented as a new map which displays the extent and location of the misclassified areas and thus in addition to the error matrix also visually presents the omission and commission errors. The analysis of polygons is however boundary dependent and can introduce positional errors of digitized boundaries to the thematic misclassification (IFEN & UNISFERE, 1997).

The point interpretation method is based on superimposing an equidistant rectangular or triangular grid over the segment. Each node is assigned the

corresponding land cover category or land use class. Pairs of nodes with equal labels obtained from the GIS and the reference data present the correct classification, while the rest present the misclassified classification (Gallego, 1995; Griffith, 1996). This method, compared to polygon interpretation is less time consuming and boundary independent (Beyeler, 1998) and therefore less expensive. However it does not enable to visualise the extent of the misclassification. For our purpose the width of the grid has been selected on the basis of the estimation of the % of classified points under nodes. Considering that the classification is performed over 10m x 10m pixels it can be for the calculation of the grid density adopted that each node covers an area of 10m x 10m. A 100m x 100m grid superimposed over the segment (3,000 m x 3,100 m) would thus have 961 nodes or cover 9.61 hectares what is ~1% of the segment area.

The results obtained by both methods are presented in classification error matrix (Tables 1 and 2).

Table 1: Error matrix of classified polygons from ground truth data and from the compiled Land Cover/Use GIS of Slovenia - state 97' of the 3,000m x 3,100m segment on location UL: x=467354; y=106000 on map 1:25,000.

LC/U GIS Hectares	Ground truth - hectares Agriculture	Forest	Built-up	Water	total	User's Accuracy %
Agriculture	**639.7**	8.5	26.3	6.7	681.2	94
Forest	64.8	**30.3**	2.3	2.7	100.1	30
Built-up	25.8	0.2	**103.5**	0.1	129.6	80
Water	3.8	0.0	0.5	**9.3**	13.6	68
Total	734.1	39.0	132.6	18.8	924.5	
						Overall acc.
Producer's accuracy %	87	78	78	49		85

User's accuracy: $(X_{ii}/\Sigma X_{ij}; j= 1,...,4) \Rightarrow$ measure of commission or inclusion error.

Producer's accuracy: $(X_{jj}/\Sigma X_{ij}; i = 1,...,4) \Rightarrow$ measure of omission or exclusion error.

Overall accuracy: $(\Sigma X_{ii}/\Sigma\Sigma X_{ij}; i=1,...,4; j=1,...,4)$.

The omission and commission classification errors are presented above and below the diagonal values, which show the correctly classified points or polygons. The omission error corresponds to nondiagonal column elements and the commission errors to the nondiagonal row elements. In addition the accuracy which is the percentage of correctly classified elements of the total number of elements for each category has been computed. The user's accuracy (measure of commission error) indicates the probability that a pixel classified into a given category actually represents that category on the ground. The producer's accuracy (measure of omission error) indicates the probability of a referenced pixel being correctly classified. In addition the overall accuracy, which is the percentage of all the correctly classified elements from the total number of elements, was calculated as well (Lillesand et al., 1994). These were all descriptive techniques. Other

analytical techniques could have been used on the same set of data. However in reality the chances of missclassification vary not only by class, but also across the map, i.e. area classified, and therefore misclassification errors are almost certainly more likely in some areas than others. As more is learned about the nature of these errors, it will be possible to produce more refined models of error calculation (USGS Center for Biologocal Informatics, 1997) that will be more suited to this kind of spatialy distributed data.

Table 2: Error matrix of classified points from ground truth data and from the Land Cover/Use GIS of Slovenia-state 97' of the 3,000 m x 3,100 m segment on location UL: x=467354; y=106000 on map 1:25,000.

LC/U GIS Hectares	Ground truth - hectares					User's
	Agriculture	Forest	Built-up	Water	total	Accuracy %
Agriculture	**654**	10	26	6	696	94
Forest	65	**42**	3	2	112	37
Built-up	27	0	**112**	0	139	81
Water	4	0	0	**10**	14	71
Total	750	52	141	18	961	
						Overall acc.
Producer's accuracy %	87	81	79	56		85

User's accuracy: $(X_{ii}/\Sigma X_{ij}; j= 1,...,4) \Rightarrow$ measure of commission or inclusion error.
Producer's accuracy: $(X_{jj}/\Sigma X_{ij}; i = 1,...,4) \Rightarrow$ measure of omission or exclusion error.
Overall accuracy: $(\Sigma X_{ii}/\Sigma\Sigma X_{ij}; i=1,...,4; j=1,...,4)$.

Results obtained by both methods have a very similar distribution of correctly classified land cover categories or land use classes as well as of omission and commission errors. The conformity of both distributions confirms that the 100m distance of the grid has been well chosen.

The 85% overall accuracy is an indicator of a good interpretation of the satellite scanned data, where the smallest pixel is of size 10m x 10m.

The main misclassification appears between the 'agriculture' and 'forest'. The transitional of grassland being overgrown by bush and forest contribute most to this misclassification. The reason for the misclassification between 'agriculture' and 'built-up' lies in the classification criteria where the areas delineated in the GIS as masks of built-up features are determined by the interpretation. The delineation criteria of subjective nature is the main reason for that misclassification.

'Water' demonstrates a strong tendency to be misclassified with vegetation. A close inspection of the underlying satellite scanned data revealed that the misclassification occurred on all those locations where the banks of rivers and water bodies were under grass, bush or trees.

Since both methods result in similar distributions of errors it has been decided to use the point method, which is less time consuming, therefore cheaper and not boundary dependant for the estimation of the quality of the end GIS, covering whole Slovenia or 2,027,245 hectares.

6 Conclusion

The Land Cover/Use GIS of Slovenia-state '97 is compiled from raster and vector layers. The first being resampled to a nominal resolution of 10m x 10m, the later digitized or vectored in scales 1: 5,000 to 1: 50,000. None of the vectored data have information on their quality what hinders the estimation and possible elimination or at least minimisation of scale dependant errors as well as the propagation of aggregated position and thematic errors. It has been decided that the produced GIS will be divided into blocks concurring with the areas covered by maps in scale 1:25,000. For all the 173 blocks that include at least 75% of Slovenia i.e. blocks on the state border with less than 25% coverage will be excluded, the 100 m equidistant point interpretation of segments of size 3km x 3.1km will be performed. Per one block 961 points from the segment will be interpreted. The quality estimation of the compiled Land Cover/Use GIS of Slovenia-state'97 will be thus based on the interpretation of 166,253 nodes of the superimposed equidistant grid over systematic distributed segments over the whole GIS.

References

[1] Berry, J.K (1987): Computer-assisted map analysis: potential and pitfalls. *Photogrametric Engineering and Remote Sensing*, **53**, 1405-1410.

[2] Beyeler, A. (1998): *Arealstatistik der Schweiz, die Bodennutzung in den Kantonen Bern, Luzern, Obwalden, Nidwalden, Gemeindeergebnisse 1979/85 und 1992/97.* Bundesamt für Statistik, Bern, 7-31.

[3] Gallego, F.J. (1995): *Sampling Frames of Square Segments,* An agricultural information system for the European Union, Joint Research Centre, EUR 16317 EN, 68.

[4] Griffith, D.A. (1996): The need for spatial statistics. Ch. 1. In S.L. Arlinghaus (Ed.), *Spatial Statistics*. Practical handbook. N.Y.: CRS Press, 1-17.

[5] IFN & UNISFERE, European Topic Centre on Land Cover, Workshop on Land Cover Applications - Need and Use, Copenhagen, 12-13 May 1997.

[6] Ivačič, M., Žvokelj, B.P., and Sever, G. (1997): *Spatial Data Quality Determination for the Elements of Physical Planning in Slovenia*, GV, **41,** 21-28.

[7] Land use statistics, description of some observation systems of the territory in Europe (1996), Working document no LAND/4; Luxembourg, 29.August 1996, 14.

[8] Leung, Y. (1998): A locational error model for spatial features, Int. J. *Geographical Information Science*, **12**, 607-620.

[9] Lillesand, T.M. and Kiefer, R.W. (1994): Remote Sensing and Image Interpretation, N.Y.: John Wiley & Sons, Inc., 611-618.

[10] Lojović, E.H., Šabić, D., and Tretjak A. (1998): Land Cover/Use GIS of Slovenia, Updating and Multitemporal Analyisis. Phare Multicounty COP'98 project documentation, Eurostat, DGIV-Land cover-use programme, 1-28.

[11] Marble, D. (1990): Geographic Information Systems, An overview. In D.J. Peuquet and F. Marble (Eds.) *Introductory readings in Geographic information systems*. Bristol: Taylor & Francis, Inc.

[12] Rapid Reports no 42/98: Territory and Climate, Land cover GIS of Slovenia, Statistical Office of the Republic of Slovenia, 9.

[13] Šabić, D., Tretjak, A., and Perdigao, V. (1995): *MARS 1994 Regional Inventory of Slovenia.* Grouind Survey, Yield Survey, Statistical Office of the Republic of Slovenia and Joint research Centre in Ispra, Contract no.: 10371-94-07 F1ED ISP SLO, 43.

[14] Wong, D. (1996): Aggregation Effects in Geo-Referenced Data. Ch. 5. In S.L. Arlinghaus (Ed.), *Spatial Statistics*. Practical handbook, N.Y.: CRS Press, 83-106.

[15] USGS Center for Biologocal Informatics (1997): Vegetation Mapping Program, 1997.