Accuracy assesment of the high resolution built-up map for Hungary

Prepared by:
G. Maucha and O. Petrik

31.07.2008
Version 1.1

CORINE Land Cover 2006 - Hungary
Project manager: Gergely Maucha

Institute of Geodesy, Cartography and Remote Sensing - FÖMI

Bosnyák tér 5.
H-1149 Budapest
Hungary

Contact:

+36 1 460 4176 Phone
+36 1 252 8282 Fax
mail: maucha.gergely@fomi.hu
http://www.fomi.hu

TABLE OF CONTENTS

1 Background 2

2 Methodology 3

2.1 The delivered dataset 3
2.2 Preparation of the target database for validation 3
2.3 Reference data for accuracy assesment 5
2.4 Validation methodology 5

3 Results of the validation 7

3.1 Quantitative results of the reference classification 7
3.2 Statistical evaluation of the validation results 7
3.3 Validation of the sealing level ..... **Error! Bookmark not defined.**

4 Conclusion 9

5 Some additional non-quantitative comments 10

6 References 11

1 BACKGROUND

Based on requirements of the EC and other users, in March 2006 EEA put forward a proposal to collaborate with the European Space Agency (ESA) and the European Commission (EC) on the implementation of a Fast Track Service Precursor on land monitoring, in line with the Communication from the Commission to the Council and the European Parliament "Global Monitoring for Environment and Security (GMES): From Concept to Reality" (COM(2005) 565 final).

The project builds on the benefits of GMES by combining the CORINE land cover (CLC) update with the production of additional high resolution data for a selected number of land cover classes such as those concerning built-up areas and forests. The shortcomings of a standard CLC update, which is deemed insufficient to meet the wide range of user needs, can be solved by the creation of complementary high resolution land cover data for a selected number of classes.

ESA has provided high resolution satellite imagery (SPOT-4 and IRS LISS-3) for the purposes of the project, in two time windows selected by the countries for the years 2006+-1. DLR has provided high quality orthorectification (IMAGE2006).

The producers of the high-resolution layers (the Service Providers) are using the same imagery as the national CLC2006 teams. This document is dedicated to the high resolution layer:

- built-up layer: per-pixel based classification of built-up and non-built-up areas (2 classes), together with an estimate of imperviousness of the area, also called degree of "soil sealing".

Built-up areas are characterized by the substitution of the original (semi)-natural cover or water surface with an artificial, often impervious, cover. This artificial cover is usually characterized by long cover duration (FAO Land Cover Classification System, 2005). Impervious surfaces of built-up areas account for 80 to 100% of the total cover.

Service providers selected by the EEA are expected to provide a per-pixel estimate of imperviousness (continuous variable from 0 to 100 percent) as index for degree of soil sealing for the whole geographic coverage. The data are to be produced in full spatial resolution, i.e. 20 m by 20 m, which provides the best possible core data for any further analysis.

The **classification accuracy per hectare** (based on a 100 m x 100 m grid) of built-up and non-built-up areas should be at least **85%**, for the **European** product.

2 METHODOLOGY

2.1 THE DELIVERED DATASET

The high-resolution soil sealing database was produced in full spatial resolution (i.e. 20x20m), based on IMAGE2006 satellite data as a two steps precedure (Sánches and Kahabka, 2008):

1. Supervised classification of the IMAGE2006 satellite data with the following visual improvement of of classification result. Resulting classes: Non-built-up / Built-up / No data. The Built-up class compises pixels that are fully or partly covered by houses, roads, mines and quarries and any other facilities, including auxiliary spaces, deliberately installed for the pursuit of human activities. Built-up area does not include any fully vegetated pixels, even if they are closely related to these activities (such as city parks and gardens), or any other unvegetated non-built-up open spaces covered with bare soil, sand, glacier, bare rocks or water.
2. Derivation of degree of soil sealing (only for the Built-up class) based on calibrated NDVI classification of IMAGE2006 satellite data. **Built-up areas are represented** therefore according to the terms of the production **by a degree of soil sealing between 1-100%.**

2.2 PREPARATION OF THE TARGET DATABASE FOR VALIDATION

According to the EEA specifications the validation of this product should be performed on a generalized dataset containing 100x100m cells. The sealing values for the 100x100m cells should be calculated as an average of the 20x20m pixel sealing values, excluding unclassifiable pixels and no data cells. **For validation purposes another definition of Built-up areas has to be used: the 100x100m cells should be categorized as Built-up or Non-built-up areas.** This should be done by **applying the 80% threshold on the sealing level value** (Table 1).

Table 1 Interpretation of the raster values of the HR built-up product

Raster value	Production class	Validation class
0	Non-sealed surface	Non-built-up area
1-79	Sealed surface	
80-100	Sealed surface	Built-up area
254	Unclassifiable pixels	
255	No data	

Received data have been prepared as follows:

1. Preparation of a 100x100m cell sealing level database by averaging the sealing level values of the 20x20m database;
2. Preparation of the 100x100m thematic database containg only the Built-up and Non-built-up classes by applying a threshold of 80% average sealing level;
3. Application of a Stratified Random Sampling on the prepared 100x100m thematic database as a target of the validation: 500 samples have been distributed inside the Built-up class (Stratum 1) to estimate the commission error, and another 2000 samples have to be distributed inside the Non-built-up class (Stratum 2) to estimate omission errors;

As seen above, there are different definitions for Built-up areas in a production and a validation term. This fact can cause several misunderstandigs, wich should be clarified. A typical issue is the classification of mines, quarries, dirt roads and railway lines. In production terms these areas are Built-up areas, because they are “installed for the pursuit of human activities”. In validation terms these are Non-Built up areas, being not impervious similarly to bare soil, sand or bare rocks.

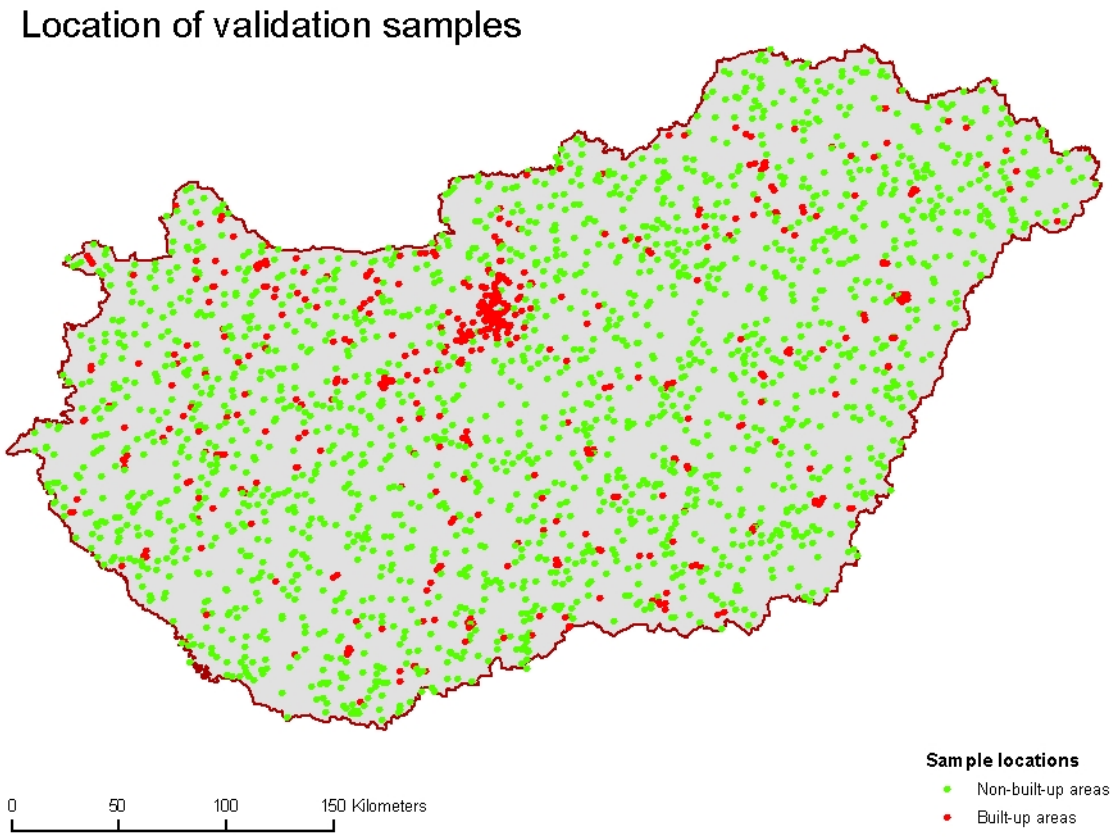

Figure 1: Location of the 500+2000 validation samples for Hungary

2.3 REFERENCE DATA FOR ACCURACY ASSESMENT

Ortho-photos taken in the years 2005/2007 have been used at 1m resolution as the base reference imagery for the accuracy assessment (Figure 2).

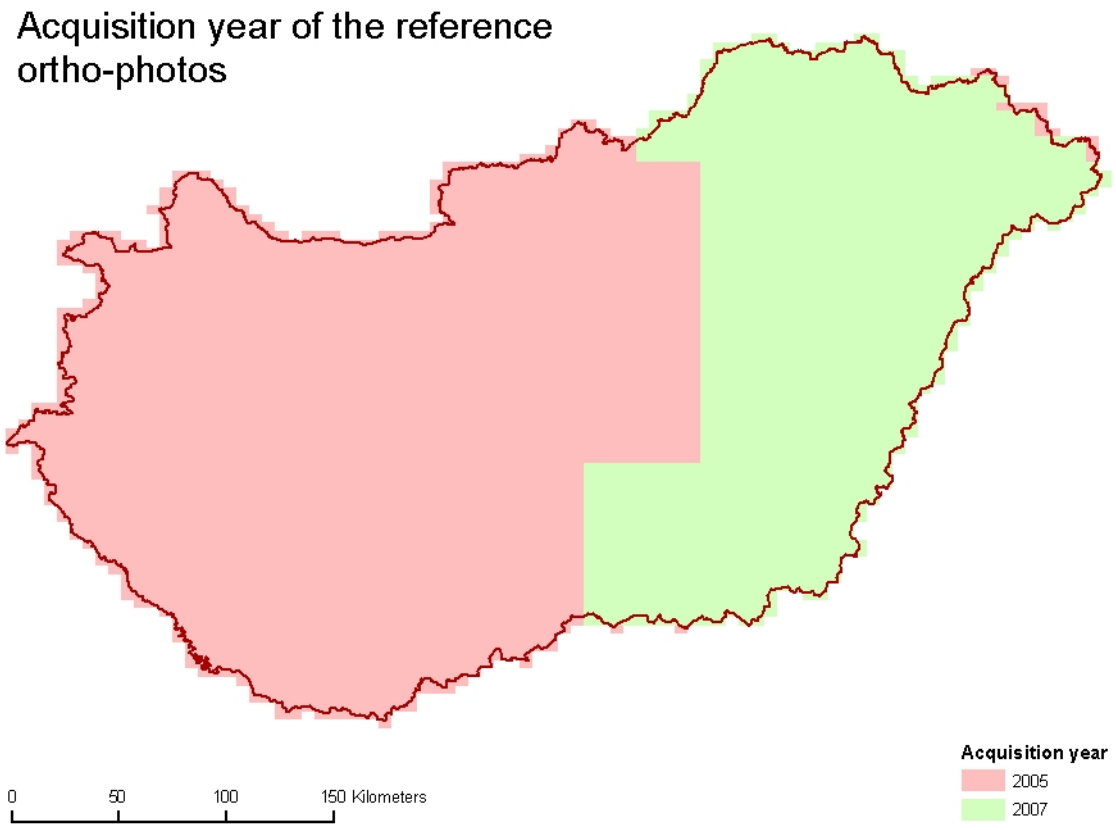

Figure 2: Acquisition year of the reference ortho-photos

1:50.000 scale scanned topographic maps (reference date: 2000 or later) and the IMAGE2006 database were available during the validation process, but these data have not been used for a decision, only for orientation purposes. In some cases Google Earth imagery has been used as a second date VHR imagery.

2.4 VALIDATION METHODOLOGY

The core of the validation methodology is the estimation of soil sealing level inside the 100x100m sample cells based on available VHR imagery. The interpretation should be "blind" in the sense that it should be carried out without having access to the HR product. All sampling cells should be evaluated and qualified, e.g. built-up/non-built-up based on an educated guess of the interpreter or other kind of visual estimation regarding the 80% sealing level.

A specific ArcView based tool has been developed by FÖMI for the purposes of the validation of soil sealing. The 100x100m sample cells have been overlaid by a 10x10 m point grid. By counting the number of impervious points out of the 100 points, the interpreter estimated the sealing degree of the sample (Figure 3).

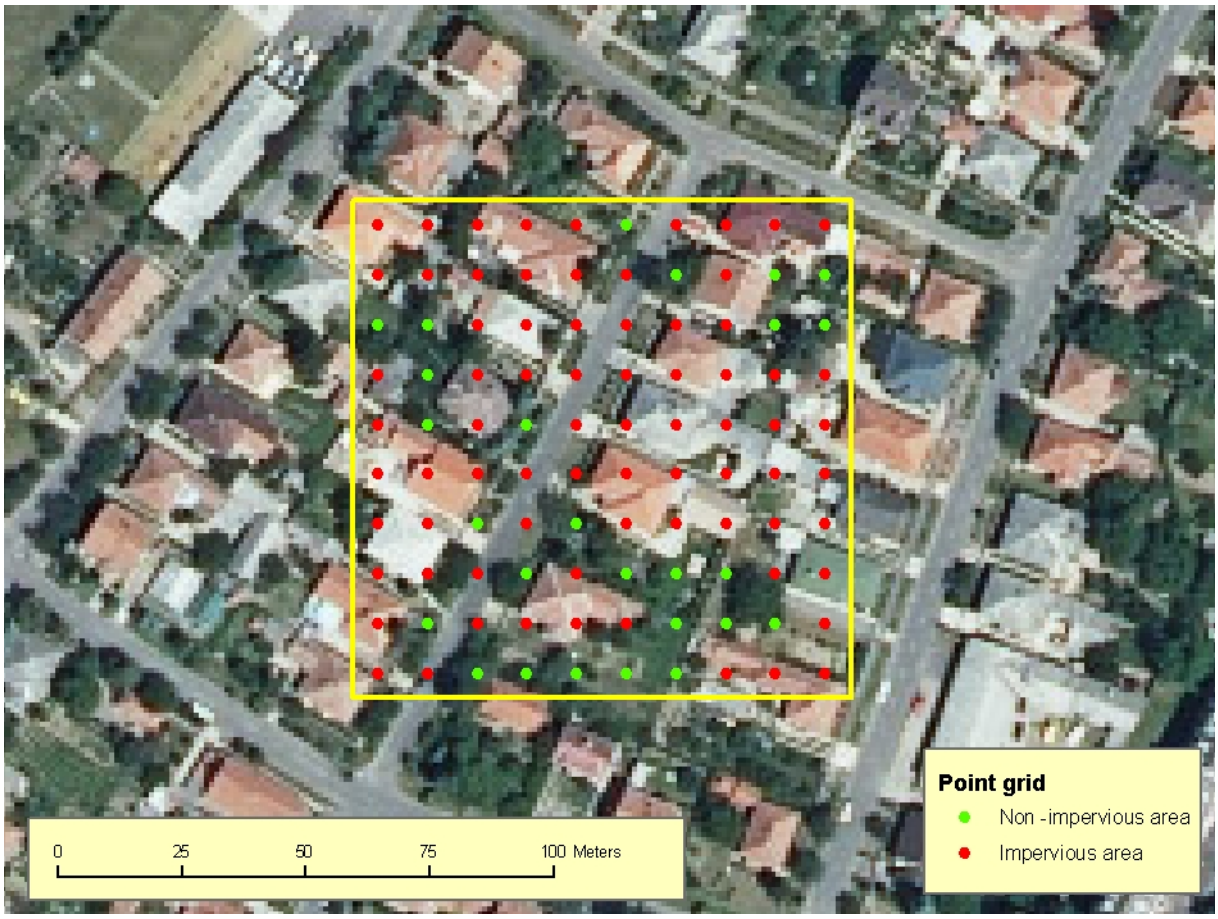

Figure 3: Visual interpretation of a sample observation (100 m x 100 m). By counting the impervious points out of the 100 points, the interpreter estimates sealing degree of the sample

The above estimation of the soil sealing degree was performed mainly in case of sample cells having soil sealing degree higher than 60%. In obvious cases the direct classification of the sample cell was applied (i.e. Non-built-up class).

3 RESULTS OF THE VALIDATION

3.1 QUANTITATIVE RESULTS OF THE REFERENCE CLASSIFICATION

During the validation process 11 of the 2500 sample cells had no corresponding reference data available (mostly because of confidentiality reasons), so altogether 2489 sample cells has been classified as Built-up / non-built-up cells, based on soil sealing degree estimated on the reference ortho-photos. In case of 600 sample cells the soil sealing value of the cell has been estimated based on the 10x10m point grid. In the case of the remaining 1889 sample cells obvious decision could have been taken without going into a detailed evaluation. Mines and quarries have been classified as non-impervious areas, but these cases have been labelled (8 cases).

Out of the 2489 valid sample cells n_1=98 samples on stratum 1 (built-up class), and n_2=1991 samples on stratum 2 (non-built-up areas) have been checked. Results are displayed in a classical error matrix are shown in Table 2.

Table 2 The error-matrix of the validation of the Built-up map in Hungary: (n_1,n_2) = (498,1991)

		Reference data			
		B	O	Sum	User's accuracy
Classified data	Built-up (B)	222	278	498	44,58%
	Other (O)	6	1 985	1 991	99.69%
	Sum	228	2 269	2489	

Because of the Stratified Sampling, the calculation of the producer's accuracy could be misleading, and therefore it is not shown.

3.2 STATISTICAL EVALUATION OF THE VALIDATION RESULTS

Statistically sound accuracy measures have been calculated with their reliabilities based on the background described in the document "Recommendations: Qantitative assesment of the high-resolution soil sealing layer" (Maucha & Büttner, 2008). The proportional area of the Built-up class in case of the 100m resolution database in Hungary is measured as P_{class} = 1,1%. No areas has been excluded from the sampling.

Results are shown in Table 3.

Table 3 Statistical results of the validation of the Hungarian HR built-up database (n_1,n_2) = (498,1991), P_{class} = 1,1%, P_{exlc}=0%, $\alpha_0=\alpha_1$=5%

Error type	Number of all samples	Number of wrong samples	Statistical results of the validation				
			p_0 (p_{min})	P_1 (p_{max})	Mean error	Confidence interval	Probability of having more than 15% error real in the database
Commission	498	278	51,84%	59,16%	55,5%	± 3,66%	100,00%
Omission	1 991	6	14,39%	53,95%	34,17%	± 19,78%	94,77%

Table 3 shows that by using 498 samples a reliable estimation can be calculated for the commission error:

- The confidence interval (including the mean commission error) is significantly over the 15% error limit and the confidence interval is modest (± 3,66%). The confidence of having more than 15% error in the database is 100%.

The 1 991 samples are not enough for a reliable estimation of the omission error:

- The 15% maximum error criteria falls into the confidence interval (p_{min} < 15%), but the confidence of having more error than 15% in the database is a bit below 95%, but the confidence interval is large (± 19,78%).

4 CONCLUSION

A quantitative accuracy assesment of the high resolution soil sealing database in Hungary has been performed by the reference interpretation of imperviousness level in sample cells based on digital ortho-photos taken in year 2005/2007. Main conclusions are as follows:

- Because of the very small proportion of Built-up area (areas with an average sealing level equal or higher than 80%) in the 100m resolution database the 1 991 sample cell can provide only estimation of the omission error with limited quality (low reliability).
- By using the 498 random sample cells inside the built-up area a good quality accuracy estimation of commission error has been provided. As the **commission error** - measured according to the EEA definitions - **is much larger than the 15% limit, the database can not be accepted**.

5 SOME ADDITIONAL NON-QUANTITATIVE COMMENTS

- The 80% threshold applied on the 100m resolution version of the soil sealing database seems to be too high to reach a meaningful built-up map from the sealing level values. In case of a significant amount of the “wrong samples” (Built-up cells = cells above 80% sealing level, but referenced as non-built-up) are measured an imperviousness level higher than 60-70%, but lower than 80%.
- The original built-up classification of the full (20m) resolution version of the database approximates much better the reality.
- False sealing values are measured mainly on surfaces with a high reflectance value (low vegetation level, sand, etc.).
- Clarification is needed how should mines, quarries, dirt roads and railway lines classified in case of the validation.

6 REFERENCES

Kopecky, M., Kahabka, H. 2008. Delivery Report Hungary

Maucha, G., Büttner, G. 2008. Recommendations for the quantitative assesment of the high resolution soil sealing layer. EEA Draft Report