Validation of Imperviousness High Resolution Layer 2006 and 2009 in Slovakia

Andrea Lajčáková[1], Konštantín Rosina[1], and Pavol Hurbánek[2]

1 Slovak Academy of Sciences, Institute of Geography, Bratislava, Slovakia

2 Catholic University in Ružomberok, Department of Geography, Slovakia

Introduction

Imperviousness High Resolution Layer (IL)

- 100 m resolution raster with 0-100% pixel values representing share of artificial impervious (built-up) surfaces;
- updated on 3-year basis (2006, 2009, and 2012 available so far);
- produced using an automatic algorithm based on calibrated NDVI;
- covers 39 countries of Europe as one of the GMES/Copernicus land monitoring services.

Previous research

- IL overestimates imperviousness degree in areas with relatively compact (urban) settlement pattern and underestimates it in areas with relatively dispersed (rural) settlement pattern (e.g. Hurbanek et al. 2010);
- countrywide or European studies (e.g. Maucha et. al 2010) of this phenomenon are rather scarce;
- usually a stratified sample is preferred, where urban areas have much higher sample rate compared to the rural area which is undersampled;
- allows for reliable estimate of commission error, while omission error is neglected, since it would require very large sample size to achieve a reasonable confidence interval (Maucha 2011).

Objectives

- Assessment of thematic accuracy of IL 2006 and 2009 in Slovakia;
- design and application of a sampling strategy that allows for estimation of error with spatial homogeneous reliability;
- comparison of alternative definitions of impervious surface when deriving the reference database;
- report on spatial and temporal variation of the accuracy indicators.

Data and methods

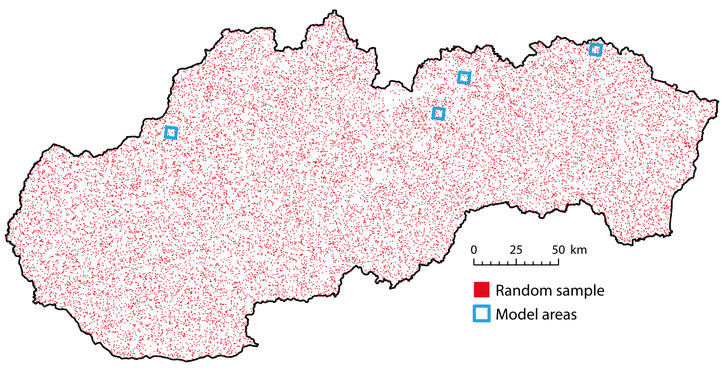
Validated datasets

- IL2006_v1 – FTSP degree of soil sealing, published in June 2009
- IL2006_v2 – revised version of 2006 data pubilshed in January 2010
- IL2006_v3 – second revision provided with 2006-2009 change layer
- IL2009

Main sampling design

- Simple (non-stratified) random sample of relatively large size, i.e. 20,000 sample plots (Fig. 1);
- each sample plot represents a single 100 × 100 m pixel of IL;
- a square lattice of 10 × 10 points is spread over each sample plot;
- aerial ortophotos are used to identify impervious points in each plot, the total number gives the reference value in the range 0-100 (see sample plot examples in Fig. 3);
- two versions of the reference database (IR) are created:
  - IR_A – more conservative definition of imperviousness, i.e. only impervious surface with allochthonous material is considered;
  - IR_B – surfaces sealed due to permanent human activity but covered with autochtonous material are considered too (i.e. dirt roads, compacted soil used for timber storage, etc.)

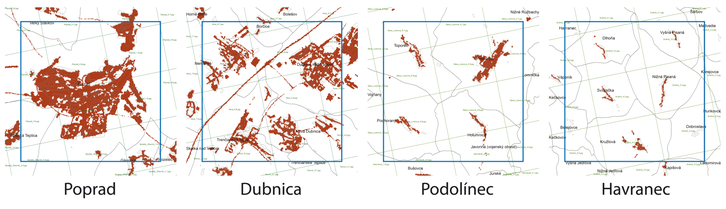
Fig. 1 Study area and employed sampling designs



Secondary sampling design

- Complete (non-sampled) reference data are produced in four model areas (6 × 6 km plots shown blue in Fig 1);
- each model area (Fig. 2) is intentionally selected to represent a different type of settlement pattern;
- each plot is mapped using a lattice of 1,200 × 1,200 points (i.e. 1,440,000 points, spaced 5 m apart).

Fig. 2 Four representatitve model areas (brown color is for positive IL values)



Evaluation methods

- Comparison of IL an IR histograms;
- correlation coefficents (Pearson's *r*, Spearman's *ρ*, Kendall's *τ*);
- error indicators – total absolute error (TAE), normalized TAE and its component errors based on error classification (Tab 1. and Fig. 3).

Measuring the errors

- We use the equations below to quantify deviation of the validated data from the reference data;
- disadvantage of TAE – it cannot be compared for territories that have different actual share of imprevious surfaces, thus we suggest using TAEN as well;

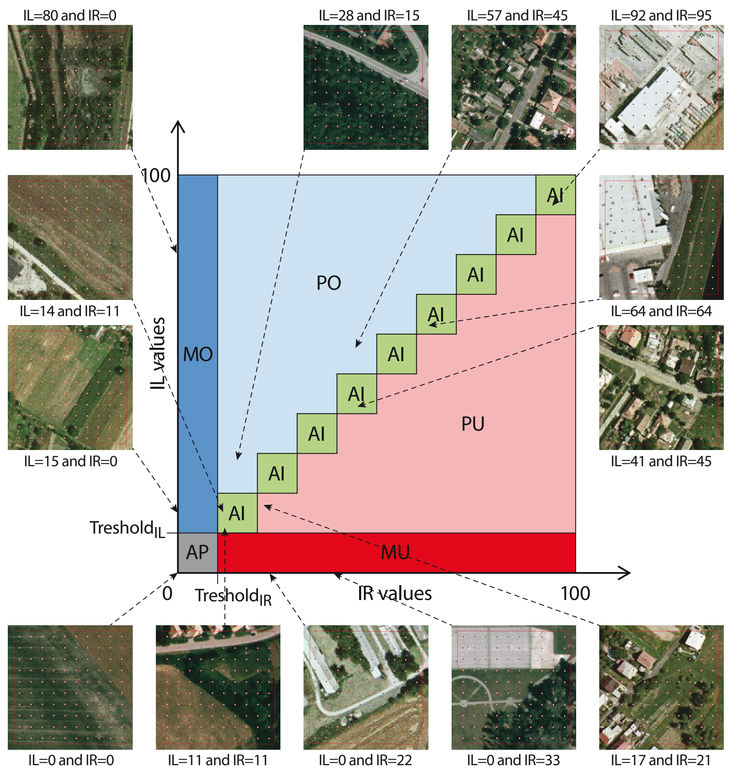
$$TAE = \sum_i^n |IL_i - IR_i| \qquad TAEN = \frac{\sum_i^n |IL_i - IR_i|}{\sum_i^n IR_i} \times 100\,(\%)$$

- it is useful to consider not only the magnitude but also the structure of the TAEN, by classfying individual pixel errors by sign and severity into partial / major over- / underestimations using a certain treshold;
- here we use 0 treshold, i.e. an error is considered major, if IL reports no impervious surface in cases where it is actually present or vice versa);
- the component parts of TAEN are suffixed by PO, PU, MO, and MU in the results section (see Tab.1 and Fig. 3 for acronym explanation).

Tab. 1 Pixel classification by type of error, value 0 is used as a treshold

IL values (IL_i)	IR values (IR_i)	Absolute difference $\lvert IL_i - IR_i \rvert$	Pixel classification
0	0	0	AP – pervious agreement
> 0	> 0	0	AI – impervious agreement
> 0	> 0	> 0	PO – partial overestimation
> 0	> 0	< 0	PU – partial underestimation
> 0	0	> 0	MO – major overestimation
0	> 0	< 0	MU – major underestimation

Fig. 3 A simplified scheme of the error types and respective sample plot examples

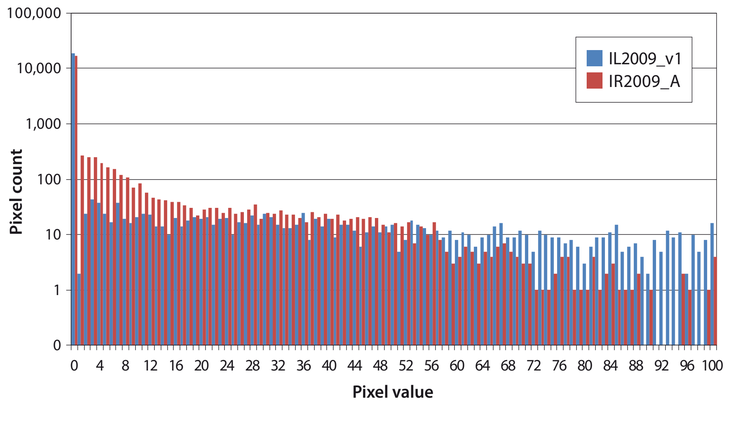


Results and discussion

Histogram comparison

- Generally, pixels having 20-55% IR value were captured by IL quite accurately;
- frequency of small imperviousness values seems to be underestimated in IL;
- frequency of large imperviousness values seems to be overestimated in IL (see Fig. 4);
- underestimation of small values is much more frequent than over-estimation of large values, which is mainly due to the fact that the small values are in reality more common than the large values.

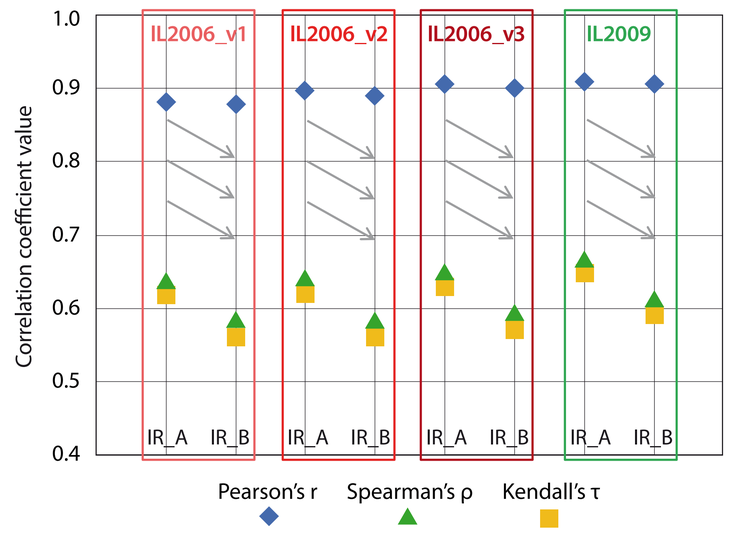
Fig. 4 Frequency distribution of IL and IR 2009 values based on the random sample



Correlation analysis

- The results obtained are consistent for the three correlation indicators employed;
- definition of IL seems to be closer to version IR_A than IR_B;
- thematic accuracy of IL2006 increased slightly from v1 to v2 and v3 in Slovakia (see Fig. 6);
- IL2009 is slightly more correlated with IR compared to all versions of IL2006.

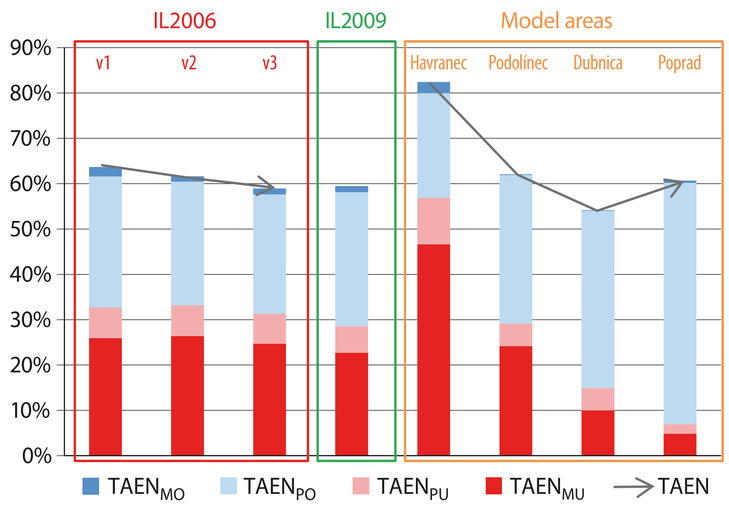
Fig. 6 Correlation coefficients between IL data and two versions of IR data (A and B)



Error indicators

- TAEN seems decreasing for 2006 data from v1 to v3 (thematic accuracy grows);
- the more rural (smaller and dispersed) are the settlements, the larger is the TAEN, especially its MU component (i.e. IL = 0 while IR > 0);
- major underestimations contribute to the overall TAEN more significantly than the partial underestimations (Fig. 7);
- for overestimations, it is the opposite – the major overestimations are much less important than the partial ones.

Fig. 7 TAEN and its structure derived from random sample as well as model areas



- Further results shown in Fig 8. might suggest that IL data are closer to IR_B definition due to smaller values of TAEN. This howeve caused by a stronger normalizing effect of the broader defined IR_B database. Smaller proportion of misclassiefied areas is attained when using IR_A instead;
- thanks to the relatively large size and non-stratified nature of the random sample, total area and share of imeprvious surfaces in Slovakia could be reliably extrapolated to the whole territory from IR data;
- obviously, the impervious area is smaller for IR_A, but it has increased by 3% in just three years.

Tab. 2 TAE, TAEN, and other statstics derived from the simple random sample

IL database	IR database	Number of underestimated pixels	Number of overestimated pixels	TAE (km²)	Proportion of misclassified areas in the total area	TAEN	Area of impervious surfaces (km²)	Proportion of impervious surfaces in the total area
IL2006_v1	IR2006_A	2,087	882	754	1.54%	65.95%	1,143	2.33%
IL2006_v1	IR2006_B	2,901	839	816	1.66%	63.68%	1,281	2.61%
IL2006_v2	IR2006_A	2,099	857	721	1.47%	63.07%	1,143	2.33%
IL2006_v2	IR2006_B	2,917	814	789	1.61%	61.62%	1,281	2.61%
IL2006_v3	IR2006_A	2,058	897	691	**1.41%**	**60.51%**	1,143	**2.33%**
IL2006_v3	IR2006_B	2,877	851	755	1.54%	58.94%	1,281	2.61%
IL2009	IR2009_A	1,971	954	706	**1.44%**	**59.89%**	1,179	**2.40%**
IL2009	IR2009_B	2,654	926	761	1.55	59.49%	1,280	2.61%

Conclusion

- Relatively small thematic accuracy, at 100 m resolution the misclassified impervious area corresponds to ca 60% of total impervious area in Slovakia, though the figures need to be assesed also in the context of other countries (similar validation is being applied in Poland and Czechia);
- the largest errors are found in rural areas with less compact, more dispersed settlement pattern; possibly due to omission of smaller and scattered objects by remote sensing methods;
- the accuracy slightly incresed from the earlier versions to the later versions, but IL should still be used with caution;
- **acknowledge its merits, be aware of its limits.**

Acknowledgement

This research was supported by the Slovak Scientific Grant Agency VEGA project no. 1/0275/13 "Production, verification and application of population and settlement spatial models based on European land monitoring services"