The distribution of selected CORINE land cover classes in different natural landscapes in Slovakia: Methodological framework and applications

Róbert PAZÚR [a], Ján OŤAHEĽ [a], Martin MARETTA [b]

Abstract

The distribution of selected CORINE land cover classes in different physical conditions was subject to modelling, analysis and evaluation in this article. In three regions with different geo-relief, the occurrence of land cover classes was analysed by using determinants commonly used in land-use models. Using three different modelling frameworks, the importance of methodological design in land-cover modelling was demonstrated. High levels of explanatory power for the factors defined here were found in landscapes of high heterogeneity. Findings derived from the statistical models highlight the importance of landscape disaggregation by natural conditions in complex land-cover or land-use models.

Keywords: land cover, land use, regression modelling, neighbourhood effect, physical properties, Slovakia

1. Introduction and research framework

If we assume that land cover (LC) at some location represents the materialisation of natural and socioeconomic conditions, then the location of each LC class relies on particular spatial relationships. According to Overmars et al. (2003), these spatial relationships can be caused by either a trend or gradient produced by the dependence of a dependent variable (e.g. LC class) upon one or more explanatory factors (independent variables), that are spatially structured, or by the interaction between the sites. It is therefore possible, to some extent, to express these spatial relationships functionally with certain related factors.

Many modelling studies, especially in the field of land use/ cover change (LUCC), consider natural and socioeconomic conditions to be the most appropriate factors. The gap between both conditional groups is often filled by derived factors, such as accessibility functions. According to Verburg et al. (2004b), all factors (also called conditions, determinants, properties or driving forces) can have biophysical, economic, social, interactive, neighbourhood or political characteristics. The influence of these factors differs widely in different conditions, however. All landscapes are historically contingent geo-systems, the structure and dynamics of which reflect continuous modifications of pre-existing systems (Demek et al., 2012).

General accessibility factors, spatial policies and neighbourhood interactions are time-varying factors and are, therefore, more important for LUCC models for the defined time periods. Identification of the distribution of LC classes is related more to the historical background of particular locations. One can compare the same territory at different times or different territories at the same time (Balej and Anděl, 2011). It is, therefore, important also to consider the research framework, which requires a strong theoretical background.

A meaningful alternative for the identification of influencing factors is the empirical investigation of factors affecting LUCC at disaggregated level / systematically stratified examples. This approach could facilitate the generalisation of a theoretical framework or the consideration of assumptions prior to the evaluation of any factor's importance. LU/LC research is struggling to develop a conceptual framework, however, that is general enough to overcome disciplinary boundaries or the complexities of comparisons beyond specific locations (Cassidy et al., 2010).

Previous studies (e.g. Ueema et al., 2008; Balej and Anděl, 2011; Pazúr et al., 2012) have shown that landscapes in common natural types have a particular landscape structure. Ueema et al. (2008), for example, showed that the spatial relationship between patches of the same LC class is lower in landscapes with a complex structure and high contrast. A physical regional boundary represents the most significant change section of natural complex characteristics, with distinct differences between the two sides of the boundary (Haibo et al., 2012). Especially in the Central European region, the transition from central planning to a market economy made the suitability of natural conditions crucial for the location of different LC classes (Balej and Anděl, 2011; Pazúr et al., 2014). The incorporation of natural properties in LU/LC studies and modelling techniques is, therefore, highly desirable.

In addition to auxiliary explanatory factors, LUCC models can be enhanced by using neighbourhood factors. Neighbourhood factors reflect forces that cannot be entirely captured by other LUCC drivers, and that are necessary to provide a statistically sound model that deals with spatial autocorrelation (Dendoncker et al., 2007).

Spatial autocorrelation is an important feature of statistical modelling with auxiliary variables and is largely influenced by spatial structures. Regression techniques employed are often cell-by-cell methods in which each pixel is treated as independent (Atkinson and Massari, 2011). Such models reflect only the global distribution and dependencies and do not reveal the self-organising nature of land development, e.g. the clustering of land uses at a local scale (Wu, 2002). To overcome these shortcomings, various

---

[a] Institute of Geography SAS, Bratislava, Slovakia; (corresponding author: R. Pazúr, e-mail: robert.pazur@savba.sk)
[b] Esprit Ltd., Banská Štiavnica, Slovakia

studies in the LUCC field integrate structural functions into regression (Braimoh and Onishi, 2007; Hengl, 2004; Hengl et al., 2007), or local estimates as in geographically-weighted regression (Brunsdon et al., 2010), or non-parametric estimates as in generalised additive models (Brown et al., 2002), neural networks (Pijanowski et al., 2005), or Bayesian models (Bogaert, 2002). Furthermore, the selection of an appropriate methodology should consider the trade-off between optimising accuracy and optimising generality (Guisan and Zimmermann, 2000).

The main aim of this paper is to investigate and properly describe the distributional agreements and differences of selected CORINE land cover classes in different natural conditions of Slovakia. For this purpose, we studied the spatial relationships of general factors that are frequently used in LUCC studies, with LC composition in three study areas that differ in their natural landscape structure.

In land use models, multiple processes can generate the same pattern, whereas the same process can generate different patterns (Verburg and Veldkamp, 2005). Therefore, land use models can be potentially improved by different parameterisations of simulation models (Lin et al., 2011). To describe these statistical relationships in the most effective way, we compare results from the application of different spatial and non-spatial models. The success or failure of the resulting models is identified in the evaluation process. From this point of view, this paper proposes a combination of methods that can be used accurately in various phases of the modelling process.

2. Materials and methods

2.1 CORINE land cover

Land cover (LC) is the materialised projection of natural spatial assets and land use, whereby natural, recreated (cultivated) and created (artificial) objects of the real landscape are identified as the physical landscape state (Oťaheľ et al., 2004). To identify and delimit the classes of LC in this paper, we used the LC layer generated under the CORINE LC (CLC) Project, often considered the standard European land cover map (Gallego et al., 2011).

The areal features represent the landscape state and consist of the 44 LC classes (31 in Slovakia), mapped to the lowest level (the third level) at a scale of 1:100 000. We used the CORINE LC 2006 data layer, which was derived from the modification of the CORINE LC 2000 data layer and by the computer-aided visual interpretation of IRS and SPOT-5 satellite images, as part of the all-European Global Monitoring for Environment and Security (GMES) Programme (Feranec and Nováček, 2009). Modifications in the data were applied if change from the previous state was greater than 5 ha or if a new patch of at least 25 ha appeared. For details of the CORINE LC methodology, see, e.g. Heymann et al. (1994), Bossard et al. (2000), or EEA (2007).

To increase representativeness, we merged the LC classes to the second classification level, which is limited to a maximum of 13 unique LC classes in the study areas. These classes are listed in Figure 1. In this paper, we particularly focused on identifying the occurrence of urban fabric, arable land, forest, and shrub or herbaceous vegetation associations, classes that are commonly present and cover most of the study areas.

2.2 Study areas

The diverse physical conditions of Slovakia played a primary role in the delimitation of study areas. We used boundaries of natural landscape types that characterize areas of native natural conditions in the present cultural landscapes. This approach is close to the mapping of the potential natural vegetation (Michalko et al., 1986).

Three study areas representing three regions of the main natural landscape types of Slovakia (lowland, basin, and highland) were selected for the purpose of this analysis (Fig. 1). Their basic statistics are listed in Table 1.

In the lowland case study (351,172 ha), biophysical conditions largely influenced the formation of the most developed area in the whole country, including the capital city of Slovakia. This is partly due to proximity to the Danube River, which was historically a major force of development throughout the whole of Europe. The river, together with climatic and soil conditions, was the basis for one of the most productive agricultural areas in this region.
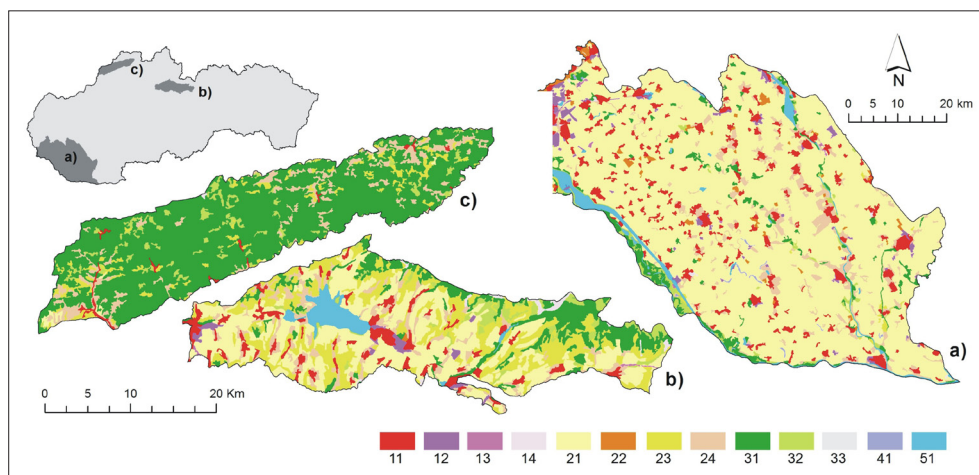

Fig. 1: LC distribution of selected study areas located in (a) lowland, (b) basin, and (c) highland landscapes. The coding of CLC classes is as follows: (11) Urban fabric; (12) Industrial, commercial and transport units; (13) Mine, dump and construction sites; (14) Artificial, non-agricultural vegetated areas; (21) Arable land; (22) Permanent crops; (23) Pastures; (24) Heterogeneous agricultural areas; (31) Forests; (32) Scrub and/or herbaceous vegetation associations; (33) Open spaces with little or no vegetation; (41) Inland wetlands; (51) Inland waters.
Source: authors´ compilation

Statistics	Case study area		
	(a) lowland	(b) basin	(c) highland
Area (ha)	351,172	59,924	50,073
Proportion (LC: %)			
Urban fabric (11)	7.40	6.72	1.10
Arable land (21)	74.71	35.91	0.83
Forest (31)	3.49	12.56	76.49
Shrubs or herb. veg. associations (32)	1.59	4.80	5.16
Other LC class	12.81	40.01	16.41
Average patch size (LC; ha)	368.12	18.44	161.06
Patch density (LC; n/ha)	0.27	0.84	0.62
Edge density (LC; m/100 ha)	23.45	51.99	40.52
Elevation (m)			
Average	116.89	709.71	663.07
Minimum	102.93	465.95	318.62
Maximum	241.80	1,158.76	1,073.83

Tab. 1: Basic statistics of selected case study areas
Source: authors´ calcultaion

Due to high bounding mountains, the basin case study area (59,924 ha) formed the centre of a specific historical, cultural and ethnic region with multifunctional land uses. Contrasting LC classes, such as urban fabric (11) and shrub and/or herbaceous vegetation associations (32), occurred there in relatively close proximity.

The highland case study (50,073 ha) represents the flysch landscape, which is a typical landscape type for this part of Europe. Hilly areas are divided here by valleys, which emphasize their morphological heterogeneity. Forests that dominated on rough and steep sites constitute a fragmented mosaic with heterogeneous agricultural areas (24), shrub and/or herbaceous vegetation associations (32) or pastures (23). Except a small patch of green urban area (part of LC class 14), the urban fabric (11) there was the only LC class characterising artificial surfaces in the CORINE LC nomenclature.

2.3 Logistic regression and determinants for the occurrence of land cover class

Land use factors that best describe land use patterns quantitatively are often selected through (logistic) regression analysis (Overmars et al., 2003). What makes logistic regression especially suitable for LUCC studies is its ability to account for binary dependent variables – the LC classes in our case.

In logistic regression, land use factors are evaluated as independent variables. A variable with the positive sign increases the probability of the occurrence of the evaluated LC class, and the negative sign has an opposite effect. If independent variables were mutually correlated and collinear, we excluded those that were less important for the modelling process in a particular case.

To identify the relevant LC determinants in the different landscapes, we only incorporate general factors that are frequently identified in LC distributional or LC change studies (Lin et. al, 2011) (Tab. 2).

Contour data were interpolated in ArcGIS 10.1 software by using the Topo to Raster tool that ensures proper preservation of the hydro-geomorphic properties of the output digital elevation model (ESRI, 2012). Soil datasets (percentages of clay, sand and silt content in top soils) were generated from the data layer of soil types. We obtained three different soil maps with identical boundaries. We assumed that the continuous nature of this subdivision would improve the forecasting potential of these datasets.

Because of complex 'cause-effect' relationships, some accessibility and socio-economic factors may be unsuitable for describing LU patterns over long time-scales (Dendoncker et al., 2007). Since we assumed that proximities to important

Dataset	Description
CORINE land cover	aggregated to level 2
Clay content (clay)	percentage generated from the soil dataset
Sand content (sand)	percentage generated from the soil dataset
Silt content (silt)	percentage generated from the soil dataset
Elevation (elevat)	digital elevation model grid with the original resolution of 20 m
Slope (slope)	digital elevation model grid with the original resolution of 20 m
Distance to towns of specific importance (city)	cost path distance; centroids of specific towns polygons
Distance to towns (town)	cost path distance; centroids of town polygons
Distance to main rivers (river)	cost path distance; rivers more than 20 m wide

Tab. 2: Datasets used. Source: authors´ calculation

rivers or urban cores represent proxies for market access and trade (Verburg, 2004b; Dendoncker et al., 2007), and therefore have an important role in accounting for urban development, we incorporated these factors into our model. For the regions located in the Slovak basins and foothills, the proximity to rivers and urban cores was especially important during the period of industrialisation. The overall accuracy of these accessibility variables was improved with the digital elevation model.

Prior to the model evaluation, all independent variables were standardised and transformed according to the following formula: $(y-min)/(max-min)$, where y is the initial value and *min* and *max* are the minimum and maximum of the original value of the independent variable (Cheng and Masser, 2003; Ondoš, 2010).

For all modelling methods we used procedures implemented in the R software environment and related packages (R Core Team, 2013).

To ensure the independence of variables, which is one of the prerequisites for the logistic regression model, we investigated the mutual correlation and co-linearity of the variables employed. For this purpose, we calculated a (Pearson) inter-correlation matrix, and validated the results by using the variance inflation factor, which measures how much the variance (square of standard error) of a coefficient is increased because of co-linearity: the higher the value of the variance inflation factor, the more serious the impact of co-linearity on the accuracy of the slope estimate (Ott et al., 2010). Because it was difficult to apply some threshold values or prejudgment rules in both quantifications, the final choice of independent variables was a subjective consideration.

2.4 Purely autoregressive model and enrichment factor

To investigate the location of the selected LC classes, we also analysed a model where the proportion of the surrounding cells occupied by the LC under study was the only independent variable. In the literature, this type of model is considered as a specific purely autoregressive (Dendoncker et al., 2007). For the neighbourhood calculations, we considered the King's case neighbourhood, which equally takes into account all eight neighbourhood cells.

In addition, for the description of the spatial interactions of selected LC classes with the whole range of LC classes, we calculated the enrichment factor. This measure describes the over- and/or under-representation of different LC classes in the particular neighbourhood of the LC class under study (Verburg et al., 2004a). Previous studies showed that the incorporation of this measure significantly improves the accuracy of LUCC models (Verburg et al., 2004a, Verburg et al., 2004b). In using the enrichment factor in the modelling framework, however, one needs to select a specific distance of the neighbourhood for each evaluated class, which may be very subjective. Incorporating multiple distances in the model causes high co-linearity. Therefore, we used the enrichment measure independently and only for informative purposes.

2.5 Spatial autocorrelation

Logistic regression is a spatially homogenous form of the generalised linear model. The per-unit effect of explanatory variables is constant across the landscape (McDonald and Urban, 2006). The spatial autocorrelation, if present, violates the statistical assumptions of independence and can lead to biased inferences (Munroe et al., 2007). Therefore, we identified the autocorrelation by analysing Moran's I index

of spatial contiguity, a standardised measure of correlation between observations in neighbouring areas (O'Sullivan and Unwin, 2012; Shortridge, 2007). The neighbourhood definition was similar to a neighbourhood defined in the purely autoregressive model. Negative values of this statistical measure are interpreted as indicating a negative spatial autocorrelation, a value near 0 indicates no spatial relationship, and positive values indicate a positive spatial autocorrelation (Shortridge, 2007).

2.6 Regression kriging

When violations of independence do occur, alternative models that account for dependence in the residuals need to be used (Keitt et al., 2002). Spatial structures, such as spatial dependency, can be described through structure functions, of which the most commonly used are correlograms, variograms and periodograms. (Overmars et al., 2003). In this study, we used the regression kriging described in detail in Hengl (2004) and Hengl et al. (2007). A simplified version of this algorithm could be described in the following steps: (1) determining the logistic model; (2) modelling the covariance structure of the logistic model residuals as a variogram; (3) interpolating logistic model residuals using simple kriging; and (4) adding the interpolated residuals surface to the logistic model surface at each prediction point. In general, the results of regression kriging might be similar to the results of logistic regression analysis in cases where spatial autocorrelation is not present.

2.7 Suitable pixel size

Prior to the analysis, all datasets were unified to the common resolution of the raster data format. The pixel size was selected by investigating the amount of the original information in the unified raster resolution. Applying the approach described in Hengl (2006) and Pazúr (in review), we finally achieved an optimal resolution of the raster cell at 80 metres.

2.8 Evaluation

The modelling procedure was performed using a split-sample approach, where the dataset was randomly divided into two groups with given proportions. The first so-called training sample (70% of all cells in the raster) was used for model calibration and parameter derivation. The quality of the model was evaluated by applying the model parameters in the second so-called test sample (30% of cells). The agreement is expressed as a proportion of the correctly allocated cells of studied LC class presence/absence.

To enhance the robustness of the evaluation we also employed other measures: the area under the curve (AUC) statistics and difference in probability. The AUC statistics expresses the model's ability to predict the probability of the occurrence of the evaluated LC class at various locations in the landscape (Braimoh and Onishi, 2007). The lower bound of the AUC statistics (0.5) expresses the entirely random assignment of modelled probabilities, while the maximum value (1.0) expresses perfect accuracy of the model. To compare the different outcomes of the different modelling approaches, we visualised the AUC values with bar-plots, an approach that was adopted from the study of Lin et al. (2011).

Statistically, more robust results of measuring the model accuracy can by achieved by quantifying the distribution and skewness of the probability values (Atkinson and Massari, 2011; Eastman et al., 2005; Sangermano et al., 2012). Therefore, we used histograms of frequencies for exploring the accuracy of probability values in the areas of occurrence

and absence of particular LC class. In this case, a perfect model would obtain the probability values of 1 in all the areas of occurrence, and the probability values of 0 in all the areas of absence (Sangermano et al., 2012).

3. Results and discussion

Initial results are presented in Table 3. Because of high mutual correlation and co-linearity, we excluded the silt content variable in all study areas from further analysis. Furthermore, for the same reason, we excluded the distance to towns or distance to towns with specific importance in the basin and highland study areas (Fig. 2).

3.1 Lowland case study

The present extension of urban land in this study area was largely determined by the proximity of the capital and existing urban areas. This spatial effect has become more noticeable since the late 1990s, when processes such as suburbanization emerged in this area (Vigašová et al., 2010; Šuška, 2012; Šveda and Križan 2012; Kopecká et al., 2014).

The spatial relationships of urban proximities were also confirmed by our models and enrichment factors. The purely autoregressive model with a simple neighbourhood variable predicted the presence of urban fabric almost perfectly (Fig. 4, Fig. 5). The same is true for the purely autoregressive predictions of all LC classes under study. Regarding the evaluation of selected factors in the logistic model, only the distance to towns and the distance to towns of specific importance partly determined the presence of the urban fabric. The negative values of these coefficients confirm the expected trend of decreasing probabilities of the occurrence of urban fabric with increasing distance from urban centres. The low explanatory power of this model, however, was confirmed by the distribution of probability values (Fig. 5), as well as by the AUC statistics (0.645).

Additionally, the high autocorrelation of model residuals (Tab. 4) illustrates that for the prediction of the urban fabric, much more complex system models are necessary.

The interpolation of logistic regression residuals with regression kriging into the test dataset, and their integration into the logistic regression model, increased the accuracy as confirmed by all assessment methods employed.

The evaluation of the urban neighbourhood enrichment clearly expresses how the existence of the capital ity attracted industrial, commercial and transport units (LC class 13, Fig. 6). Climate and historical reasons had conditioned the formation of vineyards near large urban areas, which is expressed by substantial enrichment of permanent crops (LC class 22) at a relatively small distance from urban cores.

Productive soil properties in the rural part of this region determined the large occurrence of arable land. As the logistic regression model tends to predict the location of arable land almost everywhere in the lowland, it was much more difficult to predict the absence of this LC class rather than its occurrence. This trend also reflects in the distribution of purely autoregressive model probabilities. The high proportion of presence pixels also led to a high overall accuracy, stable self-enrichment values and their low standard deviations.

The interpolation of logistic regression model residuals in the regression kriging mode had a large effect on the absence identification (Tab. 4). The proportion of correctly classified absence cells (absence model accuracy) increased in this case from 8.48% to 88.40%. This example only confirms the limitation of reporting the general percentage of correctly classified cells in cases where the proportion of presence or absence of some event (LC class in our case) is small.

Mainly due to the relatively high production potential of the lowland, none of the soil variables here was selected for the logistic regression model. The most important variable in this case is slope. Since there are only a few steeper locations in this area, interpreting the distance to the town centres and the distance to open water surfaces as important variables would be more objective. Both distance factors positively influence the location of arable land. One possible explanation for the relevance of the distance to town variable is utility

		Intercept	Elevation	Slope	Clay	Sand	Silt	City	Town	River	*Intercept	*% in neighbour.
Lowland	11	−1.29	x	x	x	x	--	−2.47	−1.33	x	−7.85	16.13
	21	0.35	x	−14.72	x	x	--	1.40	x	2.12	−7.99	15.70
	31	−3.22	x	8.51	x	x	--	x	2.02	−6.41	−8.31	16.89
	32	−4.28	x	7.10	x	x	--	x	2.97	−8.32	−8.57	17.37
Basin	11	0.06	−3.47	−6.85	x	x	--	−3.67	--	x	−7.68	15.84
	21	0.65	x	−3.04	x	x	--	−2.69	--	0.27	−7.80	15.57
	31	−6.53	8.04	6.03	x	−0.930	--	x	--	x	−7.24	14.73
	32	−6.06	5.71	1.19	−1.66	x	--	1.25	--	−0.81	−7.94	16.50
Highland	11	2.38	−11.01	−5.43	2.24	x	--	−1.13	--	−5.18	−8.08	16.91
	21	−3.11	−4.78	−5.00	x	−0.895	--	x	2.42	1.99	−8.16	16.93
	31	−1.85	2.22	3.85	x	x	--	x	--	0.74	−6.53	12.77
	32	−4.85	3.12	x	1.01	x	--	1.19	--	−0.65	−7.09	14.59

Tab. 3: Logistic regression and purely autoregressive models for selected land cover classes. Note: All presented entries for variables have the significance level at least at $p = 0.05$; * intercept and explanatory variable of purely autoregressive model; x = non-significant variable; -- variable not included
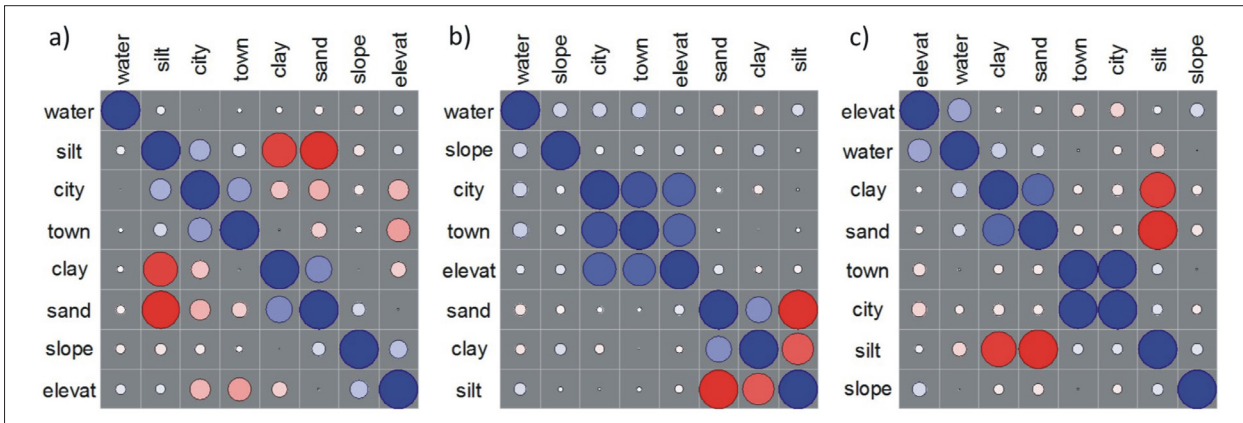Source: authors' calculations

Fig. 2: Inter-correlation matrices of independent variables in (a) lowland, (b) basin, and (c) highland landscapes: blue colour indicates positive correlation; red colour indicates negative correlation. The size of the circles indicates the strength of the correlation (r). Variables are reordered by using principal component analysis. Source: authors' calculations
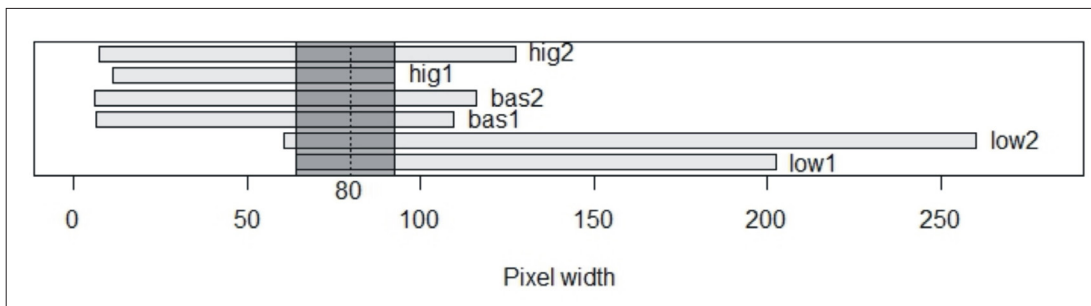

Fig. 3: Intervals of appropriate pixel resolution: (low1) LC in lowland case study; (low2) soils in lowland case study; (bas1) LC in basin case study; (bas2) soils in basin case study; (hig1) LC in highland case study; and (hig2) soils in highland case study. Source: authors' calculations

Fig. 4: AUC statistics (bars) and the proportion of correctly classified cells (star dots) for three models (LR – logistic regression, PAR – purely autoregressive, RK – regression kriging) used in three landscapes (a – lowland, b – basin, c – highland). Source: authors' calculations

maximisation and the importance of transportation costs in the overall spatial pattern of land use, as described originally by Von Thünen (von Thünen, 1966; Munroe et al., 2002). Furthermore, the importance of flood control and flood risk may be another factor contributing to the higher probabilities of the location of arable land farther from watercourses.

In this region, results that were more accurate were obtained by modelling the forested areas. Only the regression kriging model, however, resulted in valuable accuracy. Flood-plain forests were located in the neighbourhood of watercourses, mainly the Danube River. The proximity of

the river in forest neighbourhoods is also confirmed with a higher than average enrichment, especially at smaller distances. The enrichment factor also shows that areas close to the forest are frequently occupied by pastures (23) or transitional woodland/shrubs (part of the 32 LC class). Forests also cover steeper sites, and according to the logistic model, it is more probable to find forested areas here farther from the towns.

Regarding the accuracy measures, the best result for our lowland case study was obtained by the logistic prediction of shrub and/or herbaceous vegetation associations (32). The

probability distributions show, however, the highest absence prediction among all real presence samples of this LC class. Adjustments of the presence allocation with regression kriging substantially increase the probability estimations. The AUC statistics here are 0.834 for the logistic regression and 0.969 for the regression kriging. The high clustering

of this LC class was confirmed by its self-enrichment and location similarities with the forests. To some extent, it was possible to model this LC class as a function of steep sloping sites that are far from the town centres of specific importance and close to the main watercourses. The signs of significant coefficients confirmed the natural character of this LC class.
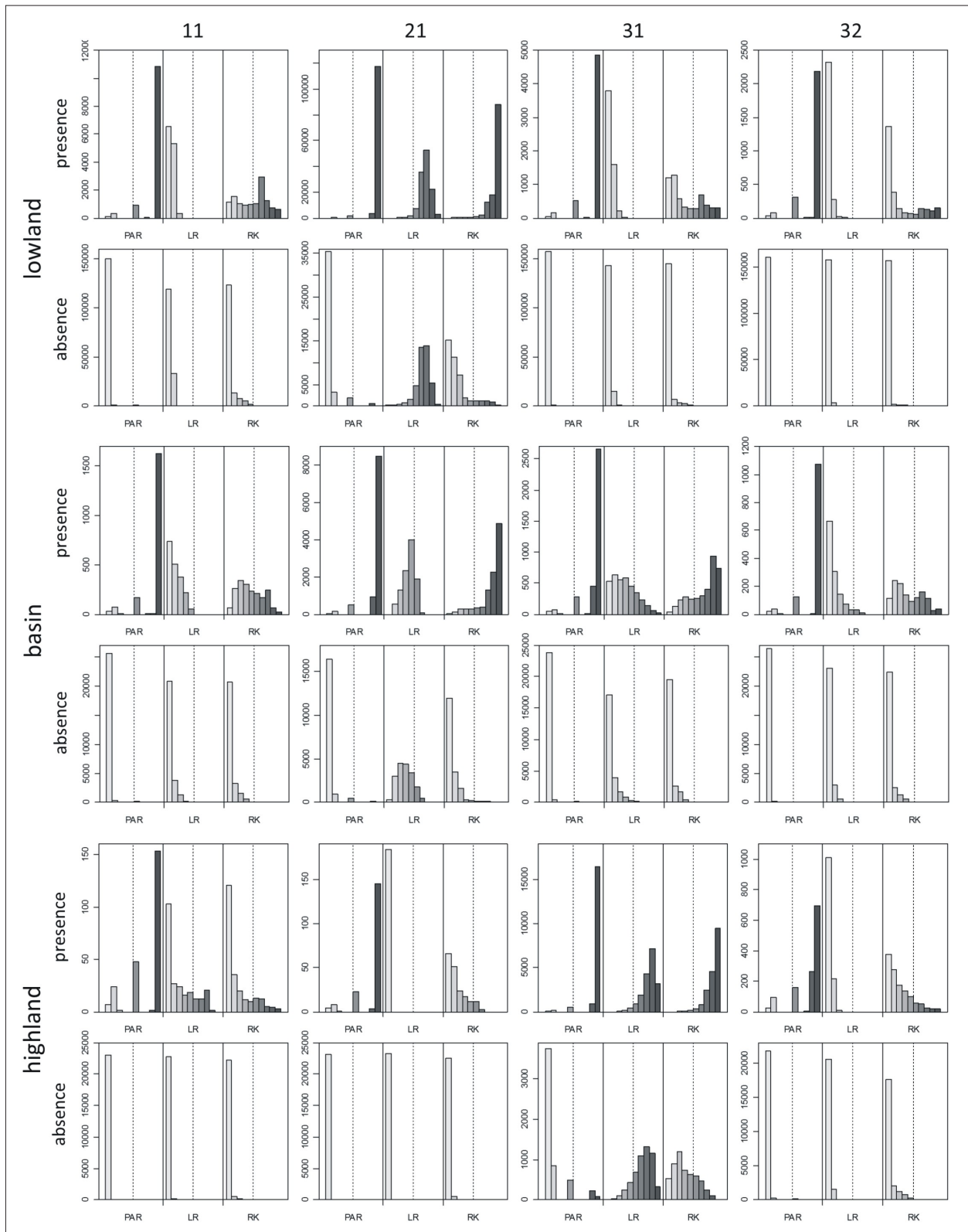

Fig. 5: Frequencies of the probability of the occurrence of selected LC classes in areas of presence and absence using different modelling approaches (LR- logistic regression, PAR – purely autoregressive, RK – regression kriging). Frequencies are binned in 10 bins within the interval 0–1. The dashed line shows the 0.5 value that represents the presence prediction threshold. Note: the y-axis is case specific. Source: authors' calculations

3.2 Basin case study

Multifunctional land use in the basin case study was confirmed by the enrichment values. High neighbourhood self-enrichments were supported with the purely autoregressive models. The AUC statistic values for the autoregressive models ranged between 0.997 and 1.0.

Acceptable probabilities also resulted from the other two modelling approaches. In the logistic model, the occurrence of the urban fabric in the basin model was determined by small slope and elevation, as well as by proximity to town centres. High model accuracy was also confirmed by the AUC statistics. The shift of the probabilities towards real values is
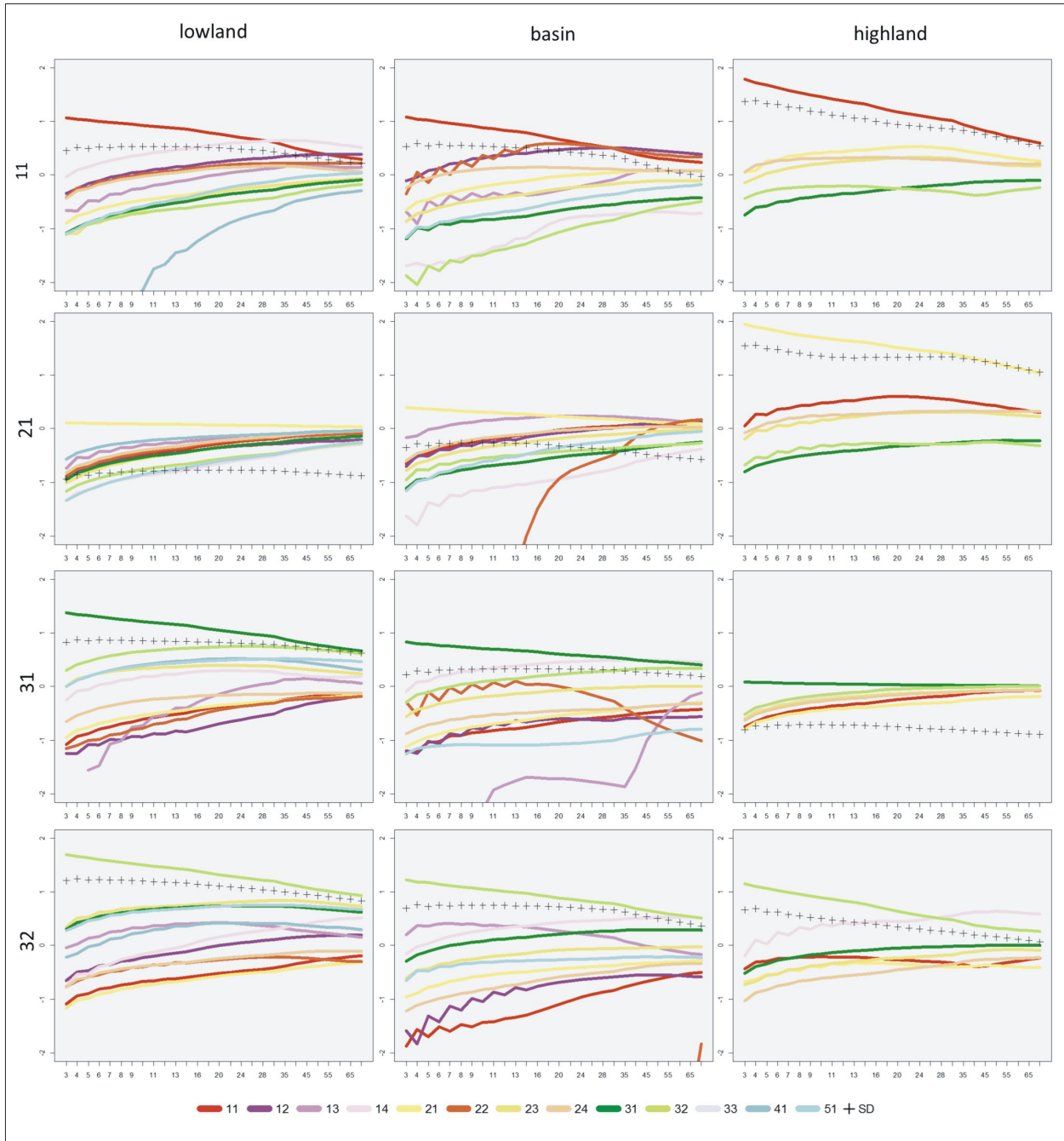

Fig. 6: Enrichment of selected LC classes and their standard deviations (SDs) at different distances. The codes are shown in Fig. 1. Note that the y-axis is at a logarithmic scale; dotted line shows standard deviation
Source: authors´ calculation

LC class	Lowland	Basin	Highland
11	0.796	0.708	0.558
21	0.822	0.781	0.657
31	0.747	0.708	0.666
32	0.718	0.716	0.687

Tab. 4: Moran I index autocorrelation values for logistic regression model residuals. Source: authors´ calculation

clear also from the distribution histograms. As demonstrated in the enrichment plot, the urban forms in this region were also positively related to the location of industrial and commercial zones (12).

Less predictable here was the absence of arable land. Similar to the urban fabric zones, some portion of the variability was explained by the selected factors, especially by elevation, distance to towns and to a smaller extent by the proximity to watercourses. The small AUC values and the close shapes of enrichment factor curves in this region, show that due to homogenous conditions arable land can occur in each locality of the study area. The global autocorrelation value (0.78) indicates the strong spatial relation of logistic model residuals. The interpolation of these residuals with the regression kriging, therefore, significantly increased the model accuracy.

In contrast to the results for arable land, the logistic model is the model for the forested area occurrence, determined by elevation, slope and silt content. Regarding the AUC statistics and probability distributions, this was the best basin logistic regression model.

In reality, forests are mainly clustered and highly enriched by themselves in northern parts of this region. Most forests in these parts occur on lands that were previously maintained for costly agricultural production. The occurrence of forests on high-reaching slopes documents that these locations were hardly suitable for agricultural activities. The residuals of the forest logistic model increased in areas where this LC class occurred. Spatial relationships are expressed with relatively high values of the Moran I global autocorrelation index. Interpolation with regression kriging improved the forest occurrence prediction from 22.13% to 73.95%, resulting in a better overall accuracy.

The occurrence of the shrub and/or herbaceous vegetation associations (32) in the basin is strongly related to forests. Positional and model similarities with forests likely indicated abandonment in the sense of transition from agricultural to forests lands (Kopecká et al., 2012; Pazúr et al., 2014). The remoteness of these locations was expressed particularly by the enrichment factor, where the urban fabric shows the lowest enrichment at almost all distances. It is also interesting that the enrichment of relatively close neighbourhoods of this LC class to garden slots is represented by the artificial, non-agricultural vegetated areas of the LC class (14). Due to the generally lower proportion of forests, the enrichment of this LC class with the other LC classes became higher only at some larger neighbourhood size.

Compared to forests, the logistic regression model for the shrub and/or herbaceous vegetation (32) was more complex. According to modelling results, we infer that the probability of the occurrence of shrub and/or herbaceous vegetation associations (32) in the basin increased in areas characterised by high elevation, slope and clay content, located farther from town centres and closer to main watercourses.

3.3 Highland case study

The urban fabric is clustered here in a few valleys across the landscape, mainly in the south-western part. Such clustering resulted in high self-enrichment and high standard deviation of the self-enrichment values. The significance of the biophysical conditions was confirmed by the logistic model, where the urban fabric could be predicted as a function of elevation, slope, distance to water, distance to town centre and by soil conditions, namely by the high clay content. Half-value logistic regression probabilities correctly predict as much as 99.06% of all urban cells, with an AUC statistic value of 0.961. These accuracies in the results are the best among all LR models. Except for soil properties, all significant variables were negatively correlated with the occurrence of urban land. Most urban fabric cells were predicted in south-western areas, where this LC class was truly concentrated. In those valleys with a low prevalence of this LC class, however, the model failed to predict its occurrence.

The high occurrence of arable land in urban neighbourhoods indicates that the two LC classes share the same conditions as described by our factors. The small overall proportion, however, makes the prediction of arable land occurrence in the mountainous landscape rather difficult. Arable land was highly enriched with itself, but the fields exhibited a considerable variability in terms of neighbourhood. Although our model correctly predicts 99.19% of all arable land occurrence, none of the predicted values exceeds even the smallest plotted probability interval of 0–0.1. Substantial adjustment of these values was provided by the interpolation of model residuals into the test sample dataset. Using the regression kriging approach, the occurrence model accuracy for this LC class increased from zero to 7.6%.

Despite the prediction errors, we obtained a good model performance with the AUC statistics. The occurrence of arable land here is significantly influenced by geo-relief (elevation, slope), soil conditions (percentage of silt content), distance to important urban centres and distance to main watercourses. Contrary to the geo-relief and soils, both accessibility variables increase the probability of the arable land occurrence in a positive way.

Elevation, slope and distance to the river were the only relevant geo-relief characteristics from the database that positively influence the location of forests. Note that such positive influence is opposite to the influence of these factors on the other LC classes under study. The results indicate that forests are inclined to occur in areas that are less suitable for intensive land use. Areas occupied by forests comprised also sites where agricultural usage had become un-economical (Demek et al., 2012). Enrichment factors and their standard deviations demonstrate that the other LC classes occur in the forest neighbourhood only rarely. A sizeable increase in forest absence prediction accuracy was obtained with the regression kriging.

The model fit for the presence of shrub and/or herbaceous vegetation associations (32) was highly inaccurate in the case of mountains (32). The general logistic model, which accounts for elevation, clay content, distances to watercourses and towns as explanatory factors, was able to predict correctly 94.73% of absence cells. Elevation, as the most significant independent variable, documented that this LC class tends to occur at higher elevations.

4. Conclusions

The composition of LC classes in diverse natural landscapes in Slovakia has been analysed in this article. In all cases, we used a similar group of factors, whose selection was based on a literature review in the field of LUCC studies. This means that we did not account for factors that might be most important for LUCC in particular cases or regions. Furthermore, we characterize the spatial distribution of LC classes in a particular region by neighbourhood enrichments.

For the modelling, we used three different regression methods. In agreement with previous modelling attempts, we found that regression kriging and purely auto-logistic regression methods have great modelling potential. Both methods can be successfully applied, for example, to improve LC change models or to improve the quality of input variables. Previous research has shown that regression kriging could efficiently improve the spatial prediction of soil variables (Hengl, 2004). One of the advantages of using purely autoregressive models is that neighbourhood-based variables can be derived using a single land-use dataset (Dendoncker et al., 2007).

Our modelling results showed that factors such as elevation and slope are strongly related to landscape heterogeneity and have, therefore, great prediction potential. If these morphometric parameters significantly vary across the landscapes, they can be used in LC modelling regardless of the landscape structure. This fact was highlighted by the agreement of coefficients that represented these factors in models of the same LC classes among different natural landscapes. A similar agreement was documented in Moravia, for example (Opršal et al., 2013).

From the LC class modelling perspective, the prediction of forests as a function of water proximities appears interesting. In lowland landscapes, forests form the surroundings of large rivers, likely for ecological (floodplain forests of high ecological value) and economic (flood risk control) reasons. In contrast, forests in highland regions, where the main watercourses form valleys with suitable living conditions, are situated mostly in remote areas with difficult access.

In using the selected variables, our predictions lead to a few broad conclusions:

• the prediction potential of more artificial LC classes increases with increasing natural heterogeneity. For LC classes of high naturalness, however, the prediction potential generally decreases with increasing heterogeneity;

• logistic regression was found to be an accurate modelling technique only in cases where models determined by the assumed factors reach high accuracies. Modelling efficiency was improved with the subsequent incorporation of autocorrelation into the model. If the aim is the prediction of the occurrence of a certain event and not knowledge about the importance of the predictors employed, then non-parametric modelling approaches, such as neural networks, may be more appropriate –especially when attempting to generalise information across data sets (Atkinson, 2004; Pijanowski et al., 2005);

• cross-sectional analysis, as a report of overall model accuracy measured in cases of rare occurrences (or rare absence), appeared to be useless. The measurement of the AUC statistics engendered more insights. Reporting model accuracy with just one value, however, tends always to be inexact (King, 1986). Respecting this caveat, we used probability histograms that were an easy-to-follow approach with high evaluation ability. On the other hand, we found such methods to be very rigid. For example, evaluation by the distribution of probability values in presence and in absence cases did not account for increasing probability trends, even if they were located in the direct neighbourhood of presence cells; and

• a common practice in LUCC research is the application and evaluation of different modelling approaches. Previous research has confirmed that case-specific testing of alternative methods is preferred to choosing a method based on arbitrary criteria or habit (Lin et al., 2011), but the role of proper evaluation appears to be underestimated. Our suggestion is to use a combination of different evaluation techniques, from which at least one will be based on a proper visualisation of the probability distribution. A good example in this regard is to follow the "good practice" recommendations described by Olofsson et al. (2014).

In this study, the models were evaluated in three different natural landscapes. We believe that these general findings about the occurrence of LC classes in different landscapes are valid and applicable for a broad range of scales and in different areas with similar natural conditions. More insights into this field may result in the extension of such models with the semantic characteristics of LC classes (Ahlqvist, 2008; Pazúr et al., 2012; Feranec et al., 2014) and/or the comparison of real and artificially generated landscapes (Kun, 2006).

Acknowledgements

Supplements

Supplementary data associated with this article can be found at http://www.geography.sav.sk/personal/pazur/analysis/distribution.html

References:

AHLQVIST, O. (2008): Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 U.S. National Land Cover Database changes. Remote Sensing of Environment, 112(3): 1226–1241.

ATKINSON, P. (2004): Spatially weighted supervised classification for remote sensing. International Journal of Applied Earth Observation and Geoinformation, 5(4): 277–291.

ATKINSON, P. M., MASSARI, R. (2011): Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy. Geomorphology, 130(1–2): 55–64.

BALEJ, M., ANDĚL, J. (2011): The role of region delimitation in a study of land cover changes: case study from the Czech Republic after 1990. Moravian Geographical Reports, 19(2): 2–17.

BOGAERT, P. (2002): Spatial prediction of categorical variables: the Bayesian maximum entropy approach. Stochastic Environmental Research and Risk Assessment, 16(6): 425–448.

BOSSARD, M., FERANEC, J., OTAHEL, J. (2000): CORINE land cover technical guide–Addendum 2000. Technical report, 40. Copenhagen, European Environment Agency.

BRAIMOH, A., ONISHI, T. (2007): Geostatistical techniques for incorporating spatial correlation into land use change models. International Journal of Applied Earth Observation and Geoinformation, 9(4): 438–446.

BROWN, D. G., GOOVAERTS, P., BURNICKI, A., LI, M. Y. (2002): Stochastic simulation of land-cover change using geostatistics and generalized additive models. Photogrammetric Engineering and Remote Sensing, 68(10): 1051–1061.

BRUNSDON, C., FOTHERINGHAM, A. S., CHARLTON, M. E. (2010): Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis, 28(4): 281–298.

CASSIDY, L., BINFORD, M., SOUTHWORTH, J., BARNES, G. (2010): Social and ecological factors and land-use land-cover diversity in two provinces in Southeast Asia. Journal of Land Use Science 5(4): 277–306.

CHENG, J., MASSER, I. (2003): Modelling urban growth patterns: a multiscale perspective. Environment and Planning A, 35(4): 679–704.

DEMEK, J., MACKOVČIN, P., SLAVÍK, P. (2012): Spatial and temporal trends in land-use changes of Central European landscapes in the last 170 years: a case study from the south-eastern part of the Czech Republic. Moravian Geographical Reports, 20(3): 2–21.

DENDONCKER, N., ROUNSEVELL, M., BOGAERT, P. (2007): Spatial analysis and modelling of land use distributions in Belgium. Computers, Environment and Urban Systems, 31: 188–205.

EASTMAN, J. R., VAN FOSSEN, M. E., SOLORZANO, L. A. (2005): Transition Potential Modelling for Land Cover Change. In: Maguire, D., Batty, M., Goodchild, M. [eds.]: GIS, Spatial Analysis and Modelling (pp. 357–385). California, ESRI Press.

EEA (2007): CLC2006 technical guidelines – European Environment Agency, Copenhagen.

ESRI (2012). ArcGIS desktop: Release 10.1. Redlands, CA: Environmental Systems Research Institute.

FERANEC, J., NOVÁČEK, J. (2009): The role of region delimitation in a study of land cover changes: case study from the Czech Republic after 1990. Moravian Geographical Reports, 17(3): 2–9.

FERANEC, J., SOLÍN, Ľ., KOPECKÁ, M., OŤAHEĽ, J., KUPKOVÁ, L., ŠTYCH, P., BIČÍK, I., KOLAŘ, J., ČERBA, O., SOUKUP, T., BRODSKÝ, L. (2014): Analysis and expert assessment of the semantic similarity between land cover classes. Progress in Physical Geography, 38(3): 301–327.

GALLEGO, F. J., BATISTA, F., ROCHA, C., MUBAREKA, S. (2011): Disaggregating population density of the European Union with CORINE land cover. International Journal of Geographical Information Science, 25(12): 2051–2069.

GUISAN, A., ZIMMERMANN, N. E. (2000): Predictive habitat distribution models in ecology. Ecological Modelling, 135(2–3): 147–186.

HAIBO, D., ZHENGFANG, W., MING, L., SHENGWEI, Z., XIANGJUN, M. (2012): Quantitative division of vegetation ecotones in Northeast China. Applied Ecology and Environmental Research, 10(3): 319–332.

HENGL, T. (2004): A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 120(1–2): 75–93.

HENGL, T. (2006): Finding the right pixel size. Computers and Geosciences, 32(9): 1283–1298.

HENGL, T., HEUVELINK, G. B. M., ROSSITER, D. G. (2007): About regression-kriging: From equations to case studies. Computers and Geosciences, 33(10): 1301–1315.

HEYMANN, Y., STEENMANS, C., CROISSILLE, G., BOSSARD, M. (1994): CORINE land cover. Technical guide. Luxembourg, Office for Official Publications of the European Communities.

KEITT, T. H., BJORNSTAD, O. N., DIXON, P. M., CITRON-POUSTY, S. (2002): Accounting for spatial pattern when modelling organism-environment interactions. Ecography, 25(5): 616–625.

KING, G. (1986): How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science, American Journal of Political Science, 30: 666–687.

KOPECKÁ, M., VATSEVA, R., FERANEC, J., OŤAHEĽ, J. STOIMENOV, A., NOVÁČEK, J., DIMITROV, V. (2012): Selected changes of arable land in Slovakia and Bulgaria during the period 1990–2006. Moravian Geographical Reports, 20(1): 43–54.

KOPECKÁ, M., VATSEVA, R., FERANEC, J., OŤAHEĽ, J., ROSINA, K. (2014): Urban land cover changes: case studies Trnava (Slovakia) and Burgas (Bulgaria). In: Bičík et al. [eds.]: Land Use/Cover Changes in Selected Regions in the World (pp. 49–55). Volume IX. International Geographical Union Commission on Land Use/Cover Change. Prague, Charles University in Prague, Faculty of Science.

KUN, Á. (2006): Generation of heterogeneous landscapes for models of population dynamics. Applied Ecology and Environmental Research, 4: 73–84.

LIN, Y. P., CHU, H. J., WU, C. F., VERBURG, P. H. (2011): Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modelling – a case study. International Journal of Geographical Information Science, 25(1): 65–87.

MCDONALD, R., URBAN, D. (2006): Spatially varying rules of landscape change: lessons from a case study. Landscape and Urban Planning, 74(1): 7–20.

MICHALKO, J., BERTA, J., MAGIC, D. (1986): Geobotanická mapa ČSSR. Veda, Bratislava.

MUNROE, D. K., SOUTHWORTH, J., TUCKER, C. M. (2002): The dynamics of land-cover change in western Honduras: exploring spatial and temporal complexity. Agricultural Economics, 27: 355–369.

MUNROE, D., NAGENDRA, H., SOUTHWORTH, J. (2007): Monitoring landscape fragmentation in an inaccessible mountain area: Celaque National Park, Western Honduras. Landscape and Urban Planning, 83(2–3): 154–167.

OLOFSSON, P., FOODY, G. M., HEROLD, M., STEHMAN, S. V., WOODCOCK, C. E., WULDER, M. A. (2014): Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148: 42–57.

ONDOŠ, S. (2010): Bunkový model súčasnej urbánnej morfogenézy na území Bratislavy, Ph.D. thesis. Bratislava, Comenius University.

OPRŠAL, Z., ŠARAPATKA, B., KLADIVO, P. (2013): Land-use changes and their relationships to selected landscape parameters in three cadastral areas in Moravia (Czech Republic). Moravian Geographical Reports, 21(1): 41–50.

O'SULLIVAN, D., UNWIN, D. J. (2010): Geographic Information Analysis. Hoboken. John Wiley & Sons, New York.

OTT, R. L., LONGNECKER, M. T. (2010): An Introduction to Statistical Methods and Data Analysis. Duxbury Press, 6th edition.

OŤAHEĽ. J., FERANEC, J., CEBECAUER, T., PRAVDA, J., HUSÁR, K. (2004): The landscape structure of the district of Skalica: assessment of changes, diversity and stability. Geographia Slovaca, 19, Bratislava, Institute of Geography.

OVERMARS, K. P, DE KONING, G. H. J., VELDKAMP, A. (2003): Spatial autocorrelation in multi-scale land use models. Ecological Modelling, 164(2–3): 257–270.

PAZÚR, R., OŤAHEĽ, J., MARETTA, M. (2012): Analýza priestorovej heterogenity tried krajinnej pokrývky v odlišných prírodných podmienkach. Geografie, 117(4): 371–394.

PAZÚR, R., LIESKOVSKÝ, J., FERANEC, J., OŤAHEĽ, J. (2014): Spatial determinants of abandonment of large-scale arable lands and managed grasslands in Slovakia during the periods of post-socialist transition and European Union accession. Applied Geography, 54: 118–128.

PAZÚR, R. (in review): Capturing the effective cell size in vector-to-raster conversion: methodological framework and its application on CORINE land cover dataset. Transactions in GIS.

PIJANOWSKI, B., PITHADIA, S., SHELLITO, B., ALEXANDRIDIS, K. (2005): Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. International Journal of Geographical Information Science, 19(2): 197–215.

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. [online] [cit. 19.04.2014]. Available at: http://www.R-project.org/

SANGERMANO, F. TOLEDANO, J. EASTMAN, J. R. (2012): Land cover change in the Bolivian Amazon and its implications for REDD+ and endemic biodiversity. Landscape Ecology, 27(4): 571–584.

SHORTRIDGE, A. (2007): Practical limits of Moran's autocorrelation index for raster class maps. Computers, Environment and Urban Systems, 31(3): 362–371.

ŠUŠKA, P. (2012): Produkcia vybudovaného prostredia v postsocialistickej Bratislave: podmienky, dynamika a územný prejav. Geografický časopis, 64(2): 155–179.

ŠVEDA, M., KRIŽAN, F. (2012): Prejavy komerčnej suburbanizácie vo vybraných odvetviach hospodárstva v zázemí Bratislavy. Ekonomický časopis, 60(5): 460–481.

UUEMAA, E., ROOSAARE, J., KANAL, A., MANDER, Ü. (2008): Spatial correlograms of soil cover as an indicator of landscape heterogeneity. Ecological Indicators, 8(6): 783–794.

VERBURG, P., DENIJS, T., RITSEMAVANECK, J., VISSER, H., DEJONG, K. (2004a): A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems, 28(6): 667–690.

VERBURG, P. H., ECK, J. R. R. V., NIJS, T. C. M. D., DIJST, M. J., SCHOT, P. (2004b): Determinants of land-use change patterns in the Netherlands. Environment and Planning B: Planning and Design, 31(1): 125–150.

VERBURG, P. H., VELDKAMP, A. (2005): Introduction to the Special Issue on Spatial modelling to explore land use dynamics. International Journal of Geographical Information Science, 19(2): 99–102.

VIGAŠOVÁ, D., PAPAJOVÁ-MAJESKÁ, Ľ., KRÍŽOVÁ, L., ŠVEDA, M. (2010): Land use changes in the suburban zones of Banská Bystrica and Zvolen (Slovakia). Moravian Geographical Reports, 18(3): 43–52.

VON THÜNEN, J. H. (1966): Von Thünen's Isolated State, Pergamon Press, Oxford, UK (P. Hall, [ed.], trans.: originally published as Der Isolierte Staat in Beziehung der Landwirtschaft und National ökonomie Wissenschaftliche Buchgesellschaft, 1875 Darmstadt).

WU, F. (2002): Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science, 16(8): 795–818.