ADVANCED SPATIAL METRICS ANALYSIS IN CELLULAR AUTOMATA LAND USE AND COVER CHANGE MODELING

ANÁLISIS AVANZADA DE MÉTRICAS ESPACIALES EN LA MODELIZACIÓN DE ALTERACIONES EN LA UTILIZACIÓN Y OCUPACIÓN DEL SUELO CON AUTÓMATAS CELULARES

ALEXANDER ZAMYATIN

PhD. Tomsk Polytechnic University, 30 Lenin Avenue, Tomsk, 634050, Russia, zamyatin@tpu.ru

PEDRO CABRAL

PhD. Instituto Superior de Estatística e Gestão de Informação, ISEGI, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 LISBOA, Portugal, pcabral@isegi.unl.pt

ABSTRACT: This paper proposes an approach for a more effective definition of cellular automata transition rules for landscape change modeling using an advanced spatial metrics analysis. This approach considers a four-stage methodology based on: (i) the search for the appropriate spatial metrics with minimal correlations; (ii) the selection of the appropriate neighborhood size; (iii) the selection of the appropriate technique for spatial metrics application; and (iv) the analysis of the contribution level of each spatial metric for joint use. The case study uses an initial set of 7 spatial metrics of which 4 are selected for modeling. Results show a better model performance when compared to modeling without any spatial metrics or with the initial set of 7 metrics.

KEYWORDS: LUCC modeling, cellular automata, transition rules, spatial metrics, geocomputation, spatial metrics contribution level

RESUMEN: Este artículo expone un abordaje que permite una definición de reglas basadas en la transición de autómatas celulares, para la modelización de alteraciones en la ocupación del suelo, aplicando un análisis avanzada de métricas espaciales. Esta considera una metodología en 4 etapas: i) búsqueda de métricas espaciales adecuadas con correlaciones mínimas; (ii) selección de una vecindad adecuada; (iii) selección de la técnica adecuada para la aplicación de las métricas espaciales; y (iv) análisis del nivel contribución de cada métrica espacial para una utilización conjunta. El caso de estudio incluye un conjunto inicial de 7 métricas espaciales de las cuales son seleccionadas 4 para la modelización. Los resultados muestran un desempeño superior tanto si se compara con una modelización sin métricas espaciales o utilizando las 7 métricas iniciales.

PALABRAS CLAVE: Modelación LUCC, autómata celular, reglas de transición, métricas espaciales, geocomputación, nivel de contribución de las métricas espaciales

1. INTRODUCTION

Land use and cover change (LUCC) modeling is one of the most promising approaches to forecast natural environmental phenomena such as forest fire spread, deforestation, soil erosion, ice and snow cover change, avalanches, etc. [1,2,3]. The LUCC models may be deterministic or stochastic [4]. Due to the stochastic nature of land cover change processes and a significant increase in computer performance, the most promising approach seems to be using stochastic models that consider spatial correlation with cellular automata (CA). A cellular automaton is a collection of cells on a regular grid, with each one having a finite number of states. Each cell evolves through discrete time steps according to transition rules based on the states of the neighboring cells. These transition rules are applied iteratively for a specified number of time steps.

Spatial metrics were developed to provide meaningful ways of measuring landscape characteristics and are usually calculated using a fixed neighborhood. There are over 100 statistical measures of spatial metrics. However, most of them are very similar and have high correlation. These metrics can be calculated with the freely available software FRAGSTATS [5]. Spatial metrics have been widely used in spatial analysis studies or in texture analysis for the classification of remotely sensed images [6,7,8].

The use of different spatial metrics in LUCC modeling has been pursued by many authors. For instance, [9] used the enrichment factor for land use pattern analysis. In [10] a set of spatial metrics for landscape structure investigation was studied. Another example reflecting the use of different sets of spatial metrics can be found in [11].

In this study, we implement LUCC modeling with CA using spatial metrics. We test whether this approach, that increases the number of factors influencing the CA transition rules, makes the LUCC modeling more adequate. The methodology considers: (i) the definition of an adequate neighborhood; (ii) a suitability map that represents the suitability of a cell changing one type of landscape to another (buffer zones, maps of distance to roads and to fresh water sources, etc.); (iii) a comprehensive set of spatial features of the studied landscape measured with selected spatial metrics.

The existence of so many spatial metrics results in the need of solving an individual subtask in CA modeling, i.e., searching for the adequate number of spatial metrics for modeling (ideally 6-7 metrics, because the calculation of each metric in a sliding window can be a computationally intensive task) and searching for the relative importance in the analyzed neighborhood. These spatial metrics should not only be sufficiently different from each other (i.e., have minimal mutual correlation and cover the definition of the different landscape characteristics), but also be efficient when used as a joint set.

The joint use of CA and a set of spatial metrics and characteristics in land cover change modeling are discussed in this paper. The approach includes a preliminary analysis of the spatial characteristics that enables their evaluation and their compatibility for joint use. The analysis is targeted at processing the minimal number of spatial metrics simplifying the calculation process and increasing modeling accuracy.

2. CA modeling and transition rules

2.1 Typical CA modeling

Cellular automata often use Markov chains to model LUCC and to forecast the thematic map $\mathbf{I'}^t$ for a certain time t in future. The procedure includes a certain number of iterations of a modeling algorithm over thematic maps $\mathbf{I}^{t2}$ and $\mathbf{I}^{t1}$ for time $t2$ and $t1$, a matrix of

transition probabilities $\mathbf{P} = [p_{ij}]$ (stochastic matrix) from the type (class) ω_i to the type ω_j, a matrix of the actual number of elements transitions within the time interval $[t2;t1]$, and a matrix of an expected number of element transitions within the time interval $[t1;t]$ assisted by CA rules as set by some f_{CA}. In this case, every CA can be viewed in the form of a matrix $\mathbf{M}_{CA} = [c_{ij}]$ with order d, where the value of the central element c_{kh} ($k = h = (2d+1)/2$, here «/» — integer division) in compliance with f_{CA} to a certain extent depends upon the value of all elements of this matrix.

The matrix $\mathbf{M}_{CA}$ will be formed by scanning a raster matrix of image $\mathbf{I}^{t1} = \{tm_{xy} = 1,2,\ldots,M; x = 1, 2,\ldots, Cn; y = 1, 2,\ldots, Rn\}$ by a *sliding window* of a size of $(2d + 1) \times (2d + 1)$ pixels and by saving the result at the current position $\{x,y\}$ of the central pixel of a sliding window as a value of pixel $c'_{kh} = f_{CA}(c_{11}, c_{12},\ldots, c_{kh},\ldots, c_{dd})$ of a new (resulting) image $\mathbf{I'}^t$ (Fig. 1). Here $\{x,y\}$ is a position of the element c_{kh} in $\mathbf{I}^{t1}$, tm_{xy} – a number of ω_i landscape type, M – a number of landscape types in the investigated area, Cn – a number of columns, and Rn – a number of rows in a thematic map.
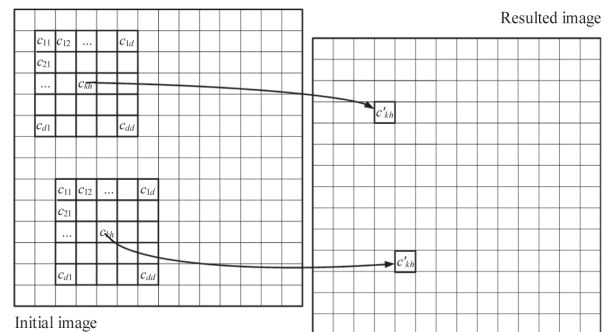

Figure 1. CA operation for collecting new values of resulted image on the basis of initial image

2.2 Definition of transition rules

Considering the aforementioned function f_{CA} for CA, the transition rules in the algorithm of land cover changes modeling should be identified on every point $\{x,y\}$ of the studied area. These are based on the probability of landscape type substitution from ω_i to ω_j as $p_{ij}^{res} \sim f_{CA}(p_{ij}^{prob}, p_{ij}^{add}, p_{ij}^{sp})$, where p_{ij}^{prob} – probabilistic characteristic, obtained on the basis of a stochastic matrix, p_{ij}^{add} – probability calculated using information from the suitability maps, p_{ij}^{sp} – probability, defined by the spatial metrics.

There are different ways of defining the functional dependence f_{CA} and finding probabilities p_{ij}^{prob}, p_{ij}^{add}, p_{ij}^{sp}. For example, the most logical way to find the joint probability p_{ij}^{res} is $p_{ij}^{\text{res}} \sim p_{ij}^{\text{prob}} \times p_{ij}^{\text{add}} \times p_{ij}^{\text{sp}}$. However, it is obvious that if the value of the probability p_{ij}^{prob}, p_{ij}^{add}, p_{ij}^{sp} is low, which is very possible in case of their multiplication, then the value p_{ij}^{res} will be very small or null. As a result, the transition from type ω_i to type ω_j will be not occur in the modeling procedure. To prevent this flaw, we may propose in practice to determine the CA transition rules in every point of the analyzed area based on the resulting transition probability of ω_i-th to ω_j-th land cover type as (1):

$$p_{ij}^{\text{res}} \sim (1 + p_{ij}^{\text{prob}}) \times (1 + p_{ij}^{\text{sp}}) \times (1 + p_{ij}^{\text{add}}) - 1 \qquad (1)$$

Following this approach (1), the multiplication of probabilities will not have the previously mentioned inconvenience, and the probabilities p_{ij}^{res} will be evenly located in the range [0,1]. It will also simplify the implementation of the modelling algorithm. The probability component p_{ij}^{prob} could be obtained as $p_{ij}^{\text{prob}} = p_{ij} \cdot n_j$ on the basis of a stochastic matrix P = $[p_{ij}]$, where n_j – number of cells with type ω_j in a CA neighborhood. The probabilistic component p_{ij}^{add}, i.e., the suitability maps, can be obtained with the help of stochastic maps built with the application of spatial analysis functions available in most GIS software. The probabilistic component p_{ij}^{sp} can also be defined by spatial metrics. One possible solution for this nontrivial task is discussed in Section 3.

3. SPATIAL METRICS APPLICATION

The main purpose of a probabilistic component p_{ij}^{sp} in CA modeling is accounting for a wide range of possible spatial characteristics of the investigated land cover. This will enable a more accurate identification of possible transition from type ω_i to type ω_j. It implies the application of spatial characteristics that precede CA modeling of land cover change. The basic concepts of this modeling approach are:the search for a limited number of spatial metrics with low correlation that account for the most important spatial characteristics of the landscape; (2) the evaluation of the size of the analyzed neighborhood in accordance with the applied data; (3) the definition of a method for accounting for the spatial characteristics in computing a probabilistic component p_{ij}^{sp}; (4) the analysis of each spatial metric

value, the definition of their optimal number, and combination for joint use in CA modeling.

The determined "optimal" finite set of metrics should be used to define a probabilistic component p_{ij}^{sp} and then to perform the final CA land cover change modeling on the basis of the expression (1). Let us consider every stage of the proposed methodology in detail:

Search for a set of metrics. At this stage, a finite number of spatial metrics should be selected. Such metrics should first of all, have minimal mutual correlation; secondly, they should cover as far as possible all the relevant spatial characteristics of the landscape. Let us define some spatial characteristic with index $l = 1, 2, …, L$ (L – total number of spatial metrics), calculated for the value $tm_{xy} \in \mathbf{I}^{tl}$ with coordinates $\{x,y\}$ in a sliding window with a neighborhood order d for a ω_k landscape type as $C_l^k = f_c(tm_{xy}, d)$, $x = 1, 2, …, Cn$; $y = 1, 2, ..., Rn$. The calculation of C_l^k for all elements of some $\mathbf{I}^{tl}$ allows one to form a matrix (defined as $\mathbf{Cm}_l^k = \{cm_{xy} = C_l^k, x = 1, 2,…, Cn; y = 1, 2,..., Rn\}$). The set of all possible metrics is rather large and will possibly have mutually dependant ones. However, most of the relevant landscape characteristics can be efficiently described using a smaller set by selecting low correlated spatial metrics. The correlation of two random spatial metrics C_l^k and $C_{l'}^k$, where $l, l' = 1, 2,…, L$, in the image $\mathbf{I}^{tl}$, can be calculated by Expression (2):

$$Cor(\mathbf{Cm}_l^k, \mathbf{Cm}_{l'}^k) = \frac{\text{cov}(\mathbf{Cm}_l^k, \mathbf{Cm}_{l'}^k)}{\sqrt{D[\mathbf{Cm}_l^k]} \cdot \sqrt{D[\mathbf{Cm}_{l'}^k]}} \qquad (2)$$

where cov – covariance, and D – variance (dispersion). Having calculated the correlation for L metrics in pairs, we can select for each C_l^k such characteristics $C_l^k{}'$ which has the least value $Cor(\mathbf{Cm}_l^k, \mathbf{Cm}_{l'}^k)$ according to (2).

This numerical evaluation allows us to solve the task of searching for a finite set of non-correlated spatial metrics. This selection can be efficiently tested regarding their capacity to reproduce the spatial characteristics of the landscape using visual expert evaluation of the computational results of spatial characteristics for ω_k landscape types, where $k = 1, 2, …, M$.

Search for a neighborhood of an appropriate size. As mentioned before, the key characteristic of any spatial metric is the size (order) of the studied neighborhood. This implies using a set of characteristics defined at the previous stage, where each characteristic can have a different size. However, there are no approaches that allow defining explicitly the most appropriate size of the studied neighborhood for a set of characteristics in CA modeling [12]. CA models are sensitive to scale and neighborhood configuration and, as a rule, the size of the neighborhood is defined by expert evaluation where all the mentioned parameters are considered [13,14,15]. The proposed approach uses all characteristics in different neighborhood sizes and evaluates the impact of all the characteristics on the modeling results.

Defining a method for accounting the spatial characteristics. The spatial characteristics for land cover change modeling with CA could be used in different ways. One of the methods for accounting such characteristics is proposed. Taking into account that there are several characteristics (L), it is reasonable to introduce them into a vector of L length, where the value of the vector components should be specified in the same range for the simplification of processing (e.g., [0;1]). Let us find the vector of "average characteristics" for each ω_k landscape type of image

$$\mathbf{I}^{t1} \; \mathbf{F}_{Cavg}^{k} = \begin{bmatrix} \overline{C}_1^{k} & \overline{C}_2^{k} \; ... & \overline{C}_L^{k} \end{bmatrix}^{\mathrm{T}}, \text{where}$$

$$\overline{C_l^{k}} = 1/Rn \times Cn \sum_{Rn \times Cn} C_l^{k}, \, l = 1,2,...,L \qquad (3)$$

In this case, the value $\overline{C_l^{k}}$ will define some average spatial statistic for ω_k landscape type. In the process of modeling and $\mathbf{I}^{t}$ image creation, we calculate for every ω_k, $k = 1, 2, ..., M,$ landscape type a vector of similar characteristics $\mathbf{F}_{Cloc}^{k} = \begin{bmatrix} C_1^{k} & C_2^{k} \; ... & C_L^{k} \end{bmatrix}^{\mathrm{T}}$, but only for some current point $\{x,y\}$ of the image $\mathbf{I}^{t1}$ in the neighborhood d. Here, the transition probability from ω_i type to ω_k type for the point $\{x,y\}$ could be defined as (4):

$$p_{ik}^{\mathrm{sp}} = |1 - d(\mathbf{F}_{Cloc}^{k}, \mathbf{F}_{Cavg}^{k})|, \qquad (4)$$

where d ($\mathbf{F}_{Cloc}^{k}$, $\mathbf{F}_{Cavg}^{k}$) – Euclidean distance between vectors.

Comparing these vectors for every landscape type in pairs, we can define how similar are both spatial characteristics of a ω_k type in the current sliding window and the average spatial characteristics on the whole raster image. This allows for the definition of the transition probability for every CA central element of the analyzed neighborhood for every ω_i landscape type, and for choosing the highest one for the ω_k landscape type.

Analysis of each metrics value in joint use. In practice, it is not easy to define the appropriate set of characteristics because of their correlation, overlapping, and other problems of joint use. To overcome this problem, the authors propose to search for the most effective set of the spatial characteristics which is described below. A series of experiments should be carried out to build a forecast map $\mathbf{I}^{t}$ based on $\mathbf{I}^{t2}$ and $\mathbf{I}^{t1}$ images by means of some unique combinations of spatial characteristics specified for every experiment. Each obtained forecast map $\mathbf{I}^{t}$ should be compared to the ground truth map $\mathbf{I}^{t}$ according to an accuracy criterion. It will allow evaluating the efficiency of the spatial characteristics combination used in every case and also the evaluation of every characteristic in the final modeling result. It should be pointed out that if the ground truth image is absent, it is still possible to evaluate the efficiency of the spatial characteristics used at this stage. In this case, the modeling should be performed in the time period [$t2$; $t1$] instead of the period [$t1$; t]; the image $\mathbf{I}^{t2}$ should be considered to be a basic image subject to transformation instead of image $\mathbf{I}^{t1}$; and a transition areas matrix should be used as the matrix of expected transition areas. In this case, the result of modeling $\mathbf{I}^{t}$ can be compared with the known image $\mathbf{I}^{t1}$ and considered a ground truth image, which enables one to evaluate the modeling results and the used set of spatial characteristics. All experiments mentioned above are aimed at selecting a unique combination of characteristics among all possible combinations. For instance, the combination $\{C_1^{k}, 0, 0,...\}$ is used in the first experiment, the combination $\{C_1^{k}, C_2^{k}, 0,...\}$ is used in the second

experiment, etc., until the very last experiment with the combination $\{C_1^k, C_2^k, …, C_L^k\}$ that includes all preselected characteristics. As a result of every experiment, we have a forecast map $\mathbf{I'}^t$. Comparing it with the ground truth map will enable us to obtain the evaluation results according to an accuracy criterion. After that, the results should be arranged in descending order according to an accuracy criterion value. Hence, after ranking each set of characteristics a value of the modeling accuracy is assigned. It allows for evaluating the value and efficiency of every characteristic in the analyzed combinations. The following methodology is recommended for the evaluation of the contribution of every characteristic to the final modeling. It is possible to evaluate a contribution level of a C_l^k characteristic

in a single h-th experiment such as $S_C^h = v_C^h \cdot \dfrac{acc^h}{M}$
, where $h = 1, 2, …, N_{expr}$, acc^h – value of accuracy criterion in an h-th experiment, M – the number of characteristics used in modeling, $v = \{0,1\}$: $v = 0$, if C_l^k is absent in a set of characteristics combination and $v = 1$ – if C_l^k is present. The contribution level evaluation of C_l^k according to the results of N experiments (i.e., integral evaluation) can be defined as (5):

$$S_C^N = \frac{1}{N} \sum_{h=1}^{N} S_C^h \qquad (5)$$

Changing successively the number of experiments $N = 1, 2, …, N_{expr}$, whose overall results define the value of every characteristic for forecast mapping accuracy, we can model the family of curves – contribution level of every characteristic based on N_{expr} experiment results. Here, it should be noted that a low number N will conform to "the most successful" combinations of characteristics, that will correspond to the most effective characteristics and ensure the most accurate modeling results. With the increase of N, a contribution level of every characteristic will be approximately the same, converging to the value (6):

$$S_C^h \approx \frac{1/N \sum_{h} acc^h}{M} \qquad (6)$$

4. RESULTS AND DISCUSSION

The proposed methodology of spatial metrics analysis and its application to CA land cover change modeling was implemented in software using C++ Builder IDE.

Let us consider in detail an application example and define the appropriate set of spatial metrics and its neighborhood size for the specific land cover change modeling case study.

In the example, the sample time series thematic maps of geoinformation software *Idrisi Kilimanjaro* – *Landuse71.rst* (image $\mathbf{I}^{t2}$), *Landuse85.rst* (image $\mathbf{I}^{t1}$) and *Landuse91.rst* (image $\mathbf{I}^t$) files were used as research data (Clark Labs – IDRISI GIS and Image Processing Software, http://www.clarklabs.org/, Office of Geographic and Environmental Information (MassGIS), Commonwealth of Massachusetts Executive Office of Environmental Affairs, http://www.mass.gov/mgis/). These maps have 9 thematic classes (High Density Residential, Low Density Residential, Industrial/Commercial, Roads/Transportation, Water, Cropland and Pasture, Forest, Wetland, Grass Surfaces) composed of 565 columns and 452 rows. Spatial resolution of the images is 60 m. The spatial pattern of the data used is quite complex which is typical for LUCC data.

From this time data series, the stochastic matrix, matrixes of transition areas, expected transition areas were calculated and the modeling was carried out on the basis of the maps for years 1971 and 1985. The map for year 1991 was used as the ground truth data.

The criteria used to evaluate the accuracy of modeling results were the *Receiver/Relative Operating Characteristic* (ROC) and *Kappa Index of Agreement* (KIA). The ROC curve is a graphical plot of the number of true positives versus the number of false positives for a binary classifier system as its discrimination threshold is varied, and it is often used to evaluate the results of a prediction. ROC of 0 means the worst agreement between an obtained map and a ground truth map; ROC of 1 means a complete agreement between an obtained map and a ground truth map. The KIA criterion is mainly used in the field of remote sensing to compare two raster images with a contingency table to find out whether their differences are due to a chance or to a real disagreement or agreement. The KIA of 1 means that two images show a perfect agreement, while the KIA of −1 means that they show a perfect and consistent agreement, and a KIA of 0 means that the two images show a random level of agreement/disagreement (i.e., there is no relationship between them). The application

of the proposed methodology is described below.

Stage 1. Selection of a set of spatial metrics. Due to the great number of applicable characteristics and according to expression (1) it is proposed to use an initial set of 7 different spatial characteristics (7)–(13). Spatial characteristics are accounted for each $\{x,y\}$ element of the raster image I^{tl} in a sliding window with $(2d + 1) \times (2d + 1)$ elements in a neighborhood for some land cover type ω_k, where d can be an order of the neighborhood:

$$C_{FO}^k = \frac{\sum_{j=1}^{n} n_j^k}{n}, \tag{7}$$

$$C_{PD}^k = \frac{m^k}{m}, \tag{8}$$

$$C_{AFM}^k = \frac{\ln(0,25 \times p_j^k)}{\ln(A^k)}, \tag{9}$$

$$C_{ADP}^k = \frac{\sum_{j=1}^{n} h_j^k}{m^k \times 2d}, \tag{10}$$

$$C_{LD}^k = \left[1 - \sum_{j=1}^{n}\left(\frac{n_j^k}{2d+1}\right)\right], \tag{11}$$

$$C_{LPS}^k = \frac{\max(m^k)}{A^k}, \tag{12}$$

$$C_{PAFRAC}^k = \frac{2}{\dfrac{\left[m\sum_{i=1}^{m}\sum_{j=1}^{n}(\ln p_j^k \cdot \ln n_j^k)\right] - \left[\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_j^k \sum_{i=1}^{m}\sum_{j=1}^{n}\ln n_j^k\right]}{\left(m\sum_{j=1}^{m}(\ln p_j^k)^2\right) - \left(\sum_{j=1}^{m}\ln p_j^k\right)^2}} \tag{13}$$

Where n is a number of elements in the neighborhood, m – a number of patches of all types, (a *patch* is a small area of one type of elements (pixels) of the landscape surrounded by the elements of another landscape type) k index determines the current ω_k landscape type, j – index determines a number of a patch), n_j^k – a number of the ω_k type elements in the neighborhood, p_j^k – a patch perimeter, h_j^k – a number of the elements to the nearest pixel of a ω_k type patch, m^k – a number of the ω_k type patches in the neighborhood,

A^k – a total area of the ω_k-type patches. For spatial characteristics, the conditional notations descriptions are: C_{FO} – frequency of occurrence (relative frequency of ω_k elements occurrence in the neighborhood), C_{PD} – level of patch density (relative frequency of specified patch), C_{AFM} – average fractal measure of perimeters (estimation of average length of patches perimeters), C_{ADP} – average distance between patches, C_{LD} – level of division (density of elements), C_{LPS} – level of patch size (patch density), C_{PAFRAC} – perimeter-area fractal dimension (evaluates patch shape complexity). These metrics were calculated for ω_k, $k = 1, 2, …, M$ using *Idrisi Kilimanjaro* raster image (file *Landuse85.rst*). Then their values were normalized into a range of 0 and 1, and prepared for accounting in p^{sp} component calculation.

When a visual assessment is needed, metrics should be presented in the form of probabilistic raster images, where each pixel represents the calculated result for the corresponding characteristic (in this case, in a 5×5 elements neighborhood size obtained for ω_k "High Density Residential Area" (Fig. 2). The visual analysis of the raster images demonstrates that every characteristic reveals some unique uncorrelated spatial features and delineates the borders between the landscape types. Obviously, characteristics C_{ADP} and C_{LD} in comparison with other characteristics revealed few spatial features of the investigated area due to the fact that the corresponding raster images (d and e in Fig. 2) contained minimum information.

Stage 2. To define the appropriate neighborhood size of spatial characteristics, a set of experiments with 3×3, 5×5, 7×7, 9×9, 15×15, and 29×29 elements in the neighborhood was carried out. The modeling was done with each characteristic from (7)–(13) separately, as well as without any spatial characteristic at all. The experiments allowed for us to prove the appropriateness of the use of spatial characteristics in CA modeling (the accuracy of the modeling results with any characteristic is significantly higher than the same results obtained without accounting for the spatial characteristics), and to find the sensitivity of every single characteristic to its neighborhood size (Fig. 3). Results show that for this type of data (landscape patterns, scale) the best accuracy results in all cases were achieved in the window size with $5 \times$

5 elements both for KIA and ROC accuracy criteria. It is obvious that the impact of spatial characteristics may significantly depend on both the neighborhood size chosen and the features (e.g., the scale) of specific landscape patterns [12,13,14].
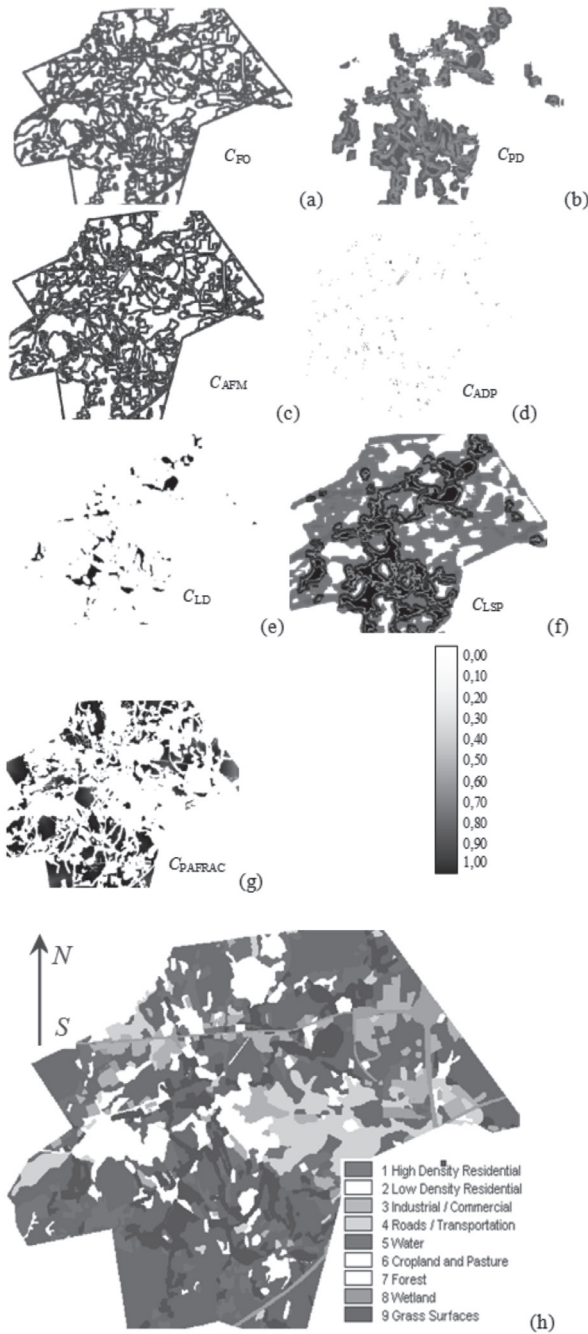
The numerical and visual evaluation of the results for different neighborhood sizes confirmed that in small neighborhood sizes, like 3 × 3, landscape spatial features are accounted to some extent only. But if we considerably increase the neighborhood size, the generalization level of spatial features becomes too high and spatial features will become hard to distinguish. Hence, to enable easy distinction of landscapes features, it is proposed to use the 5 × 5 neighborhood size for the land cover change modeling.

Figure 2. Spatial characteristics for 5 × 5 neighborhood size, obtained for ω_k-type "High Density Residential Area" (a-g), initial raster image (h) (maps without scale)
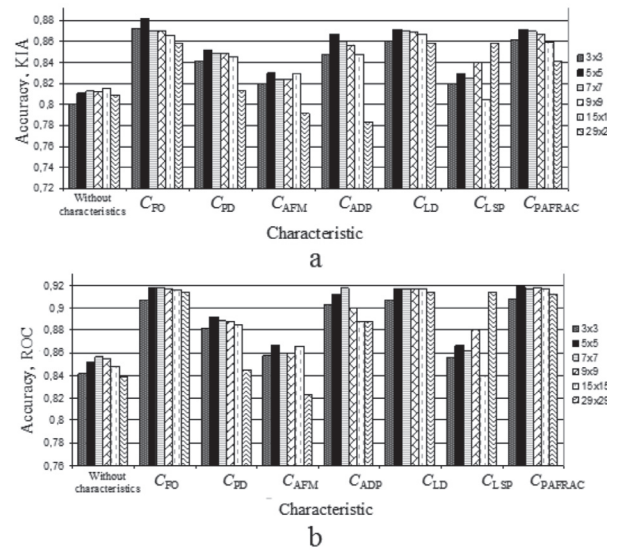

Figure 3. Accuracy of modeling results vs. spatial characteristics in different neighborhood sizes. Accuracy criteria: a) KIA; b) ROC

As an exception, the results obtained with C_{AFM} characteristic (the lowest results in comparison with other characteristics for a neighborhood of any size), and with the C_{LPS} characteristic did not show a robust correlation with a neighborhood size (results of additional experiments for C_{LPS} are depicted in Fig. 4).
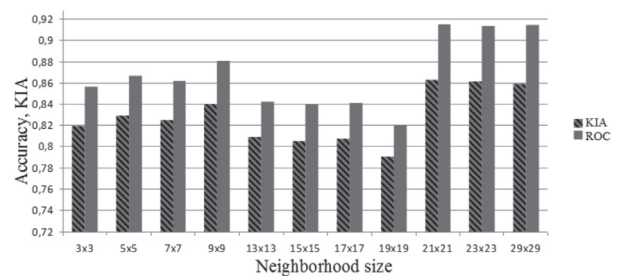

Figure 4. Accuracy of modeling results VS C_{LPS} characteristics used in different neighborhood sizes

Additional research on C_{LPS} application, performed with smaller discrete steps for neighborhood size, proved a significant sensitivity of this characteristic to the scale of the analyzed land type structure. In this case, higher and more robust results of ROC and KIA were obtained for the medium neighborhood size (5 × 5, 9 × 9 elements) and for the large neighborhood size (21 × 21, 23 × 23 and 29 × 29 elements). Considering this behavior, the C_{AFM} and C_{LPS} metrics, were omitted from this case study.

Stage 3. In accordance with the spatial characteristics application, the averaged characteristics can be presented as a vector:

$$\mathbf{F}_{Cavg}^k = \begin{bmatrix} \overline{C}_{FO}^k & \overline{C}_{PD}^k & \overline{C}_{AFM}^k & \overline{C}_{ADP}^k & \overline{C}_{LD}^k & \overline{C}_{LPS}^k & \overline{C}_{PAFRAC}^k \end{bmatrix}^T$$

The components of the vector $\mathbf{F}_{Cavg}$ are calculated for every ω_k type of the landscape by moving the sliding window throughout the raster image $\mathbf{I}^{tl}$ and calculating the mean value according to (3) of every characteristic from (7)–(13). The local characteristics of every ω_k landscape type are calculated in specific location $\{x,y\}$ using (7)–(13) in the analyzed neighborhood and can be presented as a vector:

$$\mathbf{F}_{Cloc}^k = \begin{bmatrix} C_{FO}^k & C_{PD}^k & C_{AFM}^k & C_{ADP}^k & C_{LD}^k & C_{LPS}^k & C_{PAFRAC}^k \end{bmatrix}^T$$

Furthermore, according to the proposed methodology and based on expression (4) a spatial probabilistic component p_{ij}^{sp} is defined. This component can be used in expression (1) for the final definition of CA rules for modeling. To escape the impact of p^{add} component in expression (1), we set it to zero for all $\{x,y\}$.

Stage 4. This stage is aimed at defining the appropriate set of characteristics in the case of their joint use. The overall number of experiments for 7 characteristics evaluation process is $N_{expr} = 2^7 = 128$ where all experiments are performed using a unique combination of the characteristics. Every expression resulted in a forecast map. Comparing it with the ground truth map, we may obtain the evaluation results according to the KIA and ROC criteria. After that, the results should be arranged in a descending order according to an accuracy criterion value (in this case, the KIA criterion is chosen, Fig. 5a). In Fig. 5 b, the results of the contribution level evaluation for each characteristic obtained where $N = 1, 2, …, N_{expr}$ are presented. These were obtained

according to the technique for metrics contribution level definition described by Expression (5). Results show that the first experiments have the most accurate results (Fig. 5 a, KIA ≥ 0.86, experiments no. 1-30). This means that combinations of spatial characteristics in these experiments are more "useful" for the modeling accuracy than the combinations of characteristics for experiments with the worst accuracy results (Fig. 5 a, KIA ≤ 0.86, experiments no. 50-120). The analysis of dependencies (Fig. 5 b), shows that the most precise model is obtained using the set of characteristics C_{FO}, C_{PD}, C_{LPS} and C_{PAFRAC}. This demonstrates the highest contribution level of these characteristics in the performed experiments. The smaller significance level of C_{ADP}, C_{LD} and C_{AFM} characteristics is proven by the numerical and visual evaluation (Fig. 2 d and e and Fig. 3, accordingly).
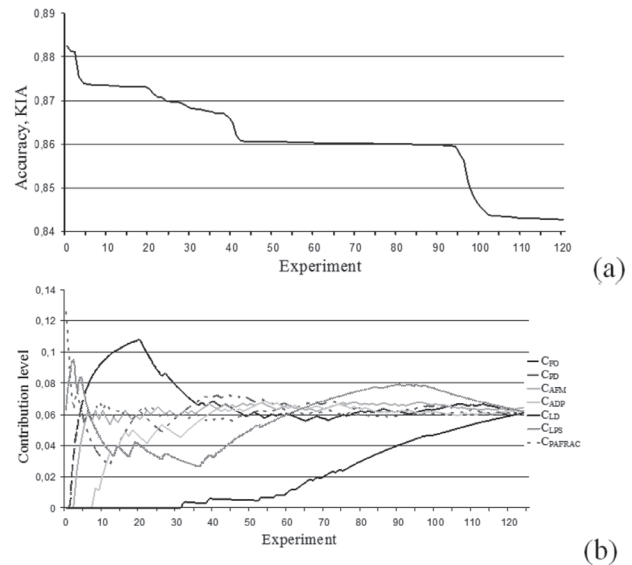

(a)

(b)

Figure 5. Results of the experiments for different combinations of spatial characteristics ranked in descending order of: a) values of KIA accuracy criterion; b) the contribution level of every characteristic

5. CONCLUSIONS

An approach to advanced spatial metrics analysis for a more effective CA transition rules definition in LUCC modeling has been proposed. This method is based on: (i) the search of the most adequate spatial metrics with a minimal mutual correlation; (ii) the definition of appropriate neighborhood size; (iii) the definition

of spatial metrics application; and (iv) the analysis of contribution level of each metric for joint use. A technique for determining the contribution level of the spatial characteristics in different sets throughout CA LUCC modeling has also been proposed. The approach was applied to time series land cover data using a set of 7 spatial characteristics. The analysis of the contribution level and the compatibility of metrics for joint use were carried out. To determine the accuracy of modeling, both ROC and KIA were used. The proposed methodology made possible the evaluation of the value of each characteristic and its compatibility using different combinations for joint use and the reduction of a number of preselected characteristics from 7 to 4 items. This lead to the simplification of the calculation process and to a modeling accuracy higher than the one obtained using any spatial characteristics or with the application of the 7 preselected spatial metrics. Results of the contribution level evaluation for each characteristic showed that the frequency of occurrence (C_{FO}), patch density (C_{PD}), level of patch size (C_{LPS}), and perimeter-area fractal dimension (C_{PAFRAC}) had the highest contribution level according to KIA. The best accuracy results for this type of data (landscape patterns, scale) were found using a neighborhood size of 5×5 elements. Due to non-robust results when using different neighborhood sizes, both fractal measure of perimeters (C_{AFM}) and level of patch size (C_{LPS}) are not recommended for use in this case study.

ACKNOWLEDGEMENTS

REFERENCES

[1]Benenson, I. and Torrens, P. Geosimulation: Automata-based Modeling of Urban Phenomena, John Wiley & Sons Ltd, Chichester, 2004.

[2] Weng, Q. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modeling, Journal of Environmental Management, 64, pp. 273-284, 2002.

[3] Cabral, P. and Zamyatin, A. Markov processes in modeling land use and land cover changes in Sintra-Cascais, Portugal, Dyna, 76 (158), pp. 191-198, 2009.

[4] Baker, W., A review of models of landscape change, Landscape Ecology, 2, pp. 111-133, 1989.

[5] Mcgarigal, K., Cushman, S. A., Neel, M. C. and Ene, E Fragstats. Spatial Pattern Analysis Program for Categorical Maps, Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html, 2002.

[6] Cushman, S. A. and. Mcgarigal, K. Patterns in the species-environment relationship depend on both scale and choice of response variables, Oikos, 105, pp. 117-124, 2004.

[7] Cushman, S.A., Mcgarigal, K. and Neel, M., Parsimony in landscape metrics: strength, universality, and consistency, Ecological Indicators, 8, pp. 691-703, 2008.

[8] Mcgarigal, K., Tagil, S. and Cushman, S.A, Surface metrics: An alternative to patch metrics for the quantification of landscape structure, Landscape Ecology, 24, pp. 433–450, 2009.

[9] Verburg, P.H., Nijsc, T., Eck, J., Visser, H. and Jong, K., A method to analyze neighborhood characteristics of land use patterns, Comput. Environ. Urban Syst., 24, pp. 354-369, 2003.

[10] Tischendorf, L., Can landscape indices predict ecological processes consistently? Landscape Ecology, 16, (3), pp. 235–254, 2001.

[11] Li, X. and Yeh, A.G.O., Data mining of cellular automata's transition rules, International Journal of Geographical Information Science, 18, pp. 723-744, 2004.

[12] Benenson, I., Warning! The scale of land-use CA is changing!, Comput. Environ. Urban Syst., 31(2), pp. 107-113, 2007.

[13] Chen, Q. and Mynett, A.E., Effects of cell size and neighbourhood type in a cellular automata based prey predator model, Simulation Modelling Practice and Theory, 11, pp. 609-625, 2003.

[14] Jantz, C.A. and Goetz S.J. Analysis of Scale Dependencies in an Urban Land Use Change Model, International Journal of Geographical Information Science, 19(2), pp. 217-241, 2005.

[15] Kocabas, V. and Dragicevic, S., Assessing cellular automata model behaviour using a sensitivity analysis approach, Comput. Environ. Urban Syst., 30 (6), pp. 921-953, 2006.