Spatial Metrics Approach to Land Cover Change Forecasting Using Cellular Automata

Alexander Zamyatin, Pavel Mikhaylov, Pedro Cabral*
Tomsk Polytechnic University, e-mail: zamyatin@tpu.ru, mpv@tpu.ru
*New University of Lisbon, Institute of Statistics and Information Management, Portugal, e-mail: pcabral@isegi.unl.pt

We represent the spatial metrics approach to the definition of cellular automata transition rules which differ from other in using of four spatial characteristics and allow increasing the modeling adequacy. The comparative analysis of the proposed approach to land cover change forecasting with the case study for Sintra-Cascais area (Portugal) is discussed.

INTRODUCTION

As a result of increasing anthropogenic impact, which causes the irreversible local and global change of nature ecosystems, the problems of the land use/cover change modeling are getting more urgent. The growth of computer facilities, hardware, algorithms and software of modern remote sensing (RS) systems increases a possibility of complex analysis and forecasting of land cover phenomena development. At present, there are some successful examples of models implementation which can be used to solve practical problems of the land cover changes forecasting (e.g. O'Sullivan D., 2000, Pontius, 2005). One of such models is the *CA_Markov* which was implemented as a program module of multifunctional raster of geoinformation system (GIS) Idrisi32 (Clark Labs, Clark University, the USA). The most complex problem while modeling with cellular automata (CA) is the definition of its transition rules. Usually, using too simple and unified CA rules (or filtration parameters) for the whole image does not allow considering many spatial characteristics of land cover types within the investigated area. Finally, it might conduct to the modeling adequacy decreasing.

The proposed spatial metrics approach to the land cover change forecasting based on four different spatial characteristics allows taking into account spatial features of the investigated area as well as CA modeling adequacy increasing.

A mathematical model tool used in the dynamic analysis and the land cover change forecasting in discrete spaces of time and states can be represented with the use of difference equation in the matrix form (e.g. Baker, 1989):

$$\boldsymbol{n}_{t+1} = \mathbf{P}\times\boldsymbol{n}_t, \qquad (1)$$

where $\boldsymbol{n}_t = [n_1,n_2,\ldots,n_m]^{\mathrm{T}}$ is a vector-column of absolute probabilities with the information to what condition among m system will transfer by the moment $(t+1)$; T is a transposition sign; P is a square matrix with rank m, where each element p_{ij} is a probability of element transition from condition Q_i to condition Q_j for the time interval from t to $(t+1)$.

Considering complexity of the ecosystems analysis as a whole, and land cover change analysis in particular, imperfections of available mathematical tools, algorithms and software of the existing models do not allow to perform effectively enough the dynamics analysis and to solve land cover forecasting problems. The majority of models are implemented as modules and subsystems in various program systems. Such models represent themselves as the research versions of software without the intensive and complex verification and analysis. The attention of researchers in the field of the land cover dynamic analysis is attracted by the most completed and frequently used models, namely: *CLUE* (*Conversion of Land Use and Its Effects*), *LUCAS* (*Land Use Change Analysis System*), *GEM* (*General Ecosystem Model*), *PLM* (*Patuxent Landscape Model*) and *UGM* (*Urban Growth Model*) (e.g. Briassoulis, 1999). Among these model implementations it is necessary to allocate UGM intended for research of urban growth and presented in the program module *CA_Markov* in GIS

Idrisi32 (e.g. Clark Labs, 2006). This model (let it be *CA_Markov* as the software) is based on the CA modeling. Because of its nature, the CA allows to consider the spatial characteristics of the investigated land cover area while modeling. To consider the spatial features and to make CA modeling much more effective it is proposed a spatial metrics approach which allows defining CA transition rules upon different spatial characteristics.

A SPATIAL METRICS APPROACH TO CA RULES DEFINITION

It is mentioned that one of the most complex problem in CA modeling is a definition of CA transition rules. For instance, some researchers (e.g. Fulong W., 2002) analyzed the necessity of components application with different probabilities to CA transition rules definition that allows considering global, local and site selection probabilities. In (Zamyatin A. & Markov N., 2005) it was proposed to define CA rules on the basis of another components within the resulting replacement probability of the i-th land type to the j-th land type in each location of the investigated area. In general, the resulting probability depends on three main components and can be represented as the follows:

$$p_{ij}^{\mathrm{res}} = f\,(p_{ij}^{\mathrm{pr}},\ p_{ij}^{\mathrm{add}}, p_{ij}^{\mathrm{sp}}), \qquad (2)$$

where p_{ij}^{pr} is a probability determined on the basis of the pure stochastic matrix $\mathbf{P}$; p_{ij}^{add} is a probability calculated on the basis of additional prior information, and p_{ij}^{sp} is a probability which is based on one spatial enrichment characteristic (e.g. Verburg, 2003). Application of several spatial characteristics for determination p_{ij}^{res} in (2) could allow considering land cover features much more effectively. But, as a rule, it is not easy due to various self-descriptiveness of these characteristics and also their essentially different range. To overcome mentioned weaknesses it is proposed an approach to probabilistic determination of CA rules, based on four spatial characteristics verified on their self-descriptiveness (e.g. Cabral, 2005). Joint application of these four characteristics to determine p_{ij}^{sp} is provided with the same range of its values, and also with the equivalent weights of coefficients.

Within the proposed approach the calculation of a land spatial characteristic is performed in the sliding window with rank d (with cardinality $(2d+1)\times(2d+1)$, at $d = 1$ the size of window 3×3). Let's consider more detailed calculation of the proposed characteristics taking into account that each component is calculated in the sliding window with the same rank d and has the normalized range $0 \div 1$.

The first characteristic indicates a *frequency of occurrence* of the k-th land type and is calculated in the following way:

$$C_{\mathrm{FO}(k,i)} = \frac{\sum_{j=1}^{m_i} n_{k,i,j}}{n_i}, \qquad (3)$$

where $n_{k,i,j}$ is the number of the k-th land type cells of the j-th sector in the i-th neighborhood, n_i is the number of cells in the neighborhood, m_i is the number of all sectors in the neighborhood. The sector here is a set of cells with one land type surrounded by cells of another land types. If the neighborhood consists of single land type, $C_{\mathrm{FO}} = 1,0$.

The second characteristic indicates a *degree fragmentation* of the k-th land type and is calculated in the following way:

$$C_{\mathrm{DF}(k,i)} = \frac{m_{k,i}}{m_i}, \qquad (4)$$

where $m_{k,i}$ is the number of the k-th land type sectors in the i-th neighborhood. For instance, if the neighborhood contains two sectors with different land types, $C_{\mathrm{DF}} = 0,5$.

The third characteristic indicates an *average fractal determination* of a k-th land type perimeters and is calculated as the following:

$$C_{\mathrm{AFD}(k,i)} = \frac{\sum_{j=1}^{m_i} \frac{2\ln(0{,}25 \times p_{k,i,j})}{\ln(n_{k,i,j})}}{m_i} - 1, \tag{5}$$

where $p_{k,i,j}$ is the j-th land type perimeter of the k-th sector in the i-th neighborhood. If all the k-th land type sectors are elementary (for example, a square), $C_{\mathrm{AFD}} = 0$.

The fourth characteristic indicates the *average distance* between the k-th land type sectors and is calculated in the following way:

$$C_{\mathrm{AD}(k,i)} = \frac{\sum_{j=1}^{m_i} h_{k,i,j}}{m_{k,i} \times 2d}, \tag{6}$$

where $h_{k,i,j}$ is the cell number from the j-th sector of the k-th land type up to the nearest j-th sector of the k-th land type in the i-th neighborhood; d is a rank of the sliding window. If sectors of one type are distributed remotely, $C_{\mathrm{AD}} = 1{,}0$.

The final replacement probabilities according to expressions (3)-(6) for the chosen land type for each i-th cell is proposed to calculate with the use of two kinds of characteristics. The first one is calculated for the current neighborhood (i.e. local characteristics); the second is for the whole image (i.e. global characteristics).

Local characteristics for each land cover types are calculated in the current neighborhood with the use of expressions (3)–(6) and can be represented in the vector form:

$$\mathbf{F}_{\mathbf{C}\,loc(k)} = \begin{bmatrix} C_{\mathrm{FO}(k,i)} \\ C_{\mathrm{DF}(k,i)} \\ C_{\mathrm{AFD}(k,i)} \\ C_{\mathrm{AD}(k,i)} \end{bmatrix}, \tag{7}$$

where k is a land cover type; i is the current neighborhood of the image cells.

Global characteristics also can be represented in the vector form:

$$\mathbf{F}_{\mathbf{C}\,gl(k)} = \begin{bmatrix} \overline{C}_{\mathrm{FO}(k)} \\ \overline{C}_{\mathrm{DF}(k)} \\ \overline{C}_{\mathrm{AFD}(k)} \\ \overline{C}_{\mathrm{AD}(k)} \end{bmatrix}, \tag{8}$$

where each component is calculated for the k-th land cover type by moving the sliding window and calculating the average value for each of characteristics (3)–(6).

Comparing values of local and global characteristics from (7) and (8) for various land cover types, it is possible to determine how much values of spatial characteristics in the sliding window are similar to the value of spatial characteristics in the whole investigated area. It allows revealing the probability of the neighborhood central cell to the k-th land type. Taking into account mentioned above the spatial probabilistic component can be calculated as:

$$p^{sp} = |1 - d(\mathbf{F}_{\mathbf{C}\ loc\ (k)}, \mathbf{F}_{\mathbf{C}\ gl\ (k)})|, \qquad (9)$$

where $d\ (\mathbf{F}_{\mathbf{C}\ loc\ (k)}, \mathbf{F}_{\mathbf{C}\ gl\ (k)})$ is Euclidean distance between (7) and (8).

The use of probabilities p_{ij}^{sp} from (9) based on expression (2) allows to consider much more spatial land features and adequately define CA transition rules in each cell of the investigated area while modeling.

LAND COVER CHANGE FORECASTING CASE STUDY

Investigated area and data use. The dynamic analysis and forecasting were carried out for Sintra-Cascais area with the square about 416 км^2 (Fig. 1), located in coastal territory to the northwest from Lisbon (Portugal). This territory is characterized by a high level of population density and its essential growth dynamics (from 1990 to 2000 growth was about 30 %). That has essentially affected the land use/cover classes and life conditions of people who lived there.

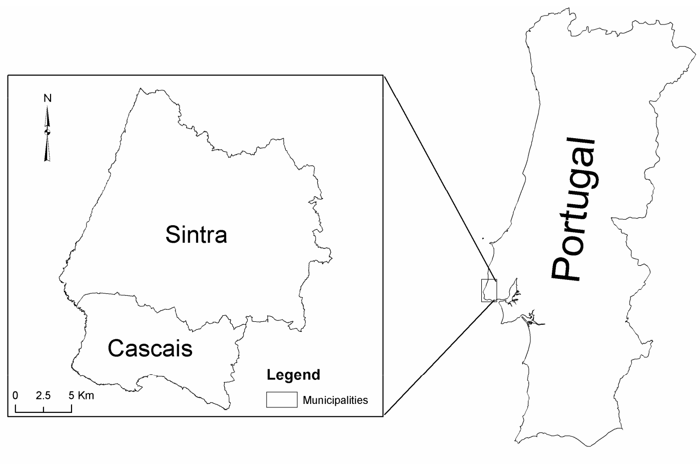


Figure 1: The study area (Sintra-Cascais).

For the dynamics analysis and forecasting, time series remotely sensed images and additional cadastre data are usually used. In this case it is possible to use land cover maps for time moments (t - 1) and t for modeling parameters calculation. The modeling procedure is carried out on the basis of obtained parameters and the land cover map for time moment t. Moreover, the land cover map was used for the moment (t + 1) of the estimation of modeling results as well as the adequacy of the model based on the proposed approach.

For the interpretation and modeling land cover change, the different time series remote sensing images of the Landsat TM (obtained 02.08.1994) and Landsat ETM+ (obtained 04.08.2001 and 03.14.1989) were used as initial data. This allows to carry out modeling with the use of the land cover maps for 1989 and 1994, and to estimate modeling performance with the use of the land cover map for 2001.

Using available ground truth observations, learning samples have been constructed for each remotely sensed image of such land types both for populated and unpopulated areas. Remotely sensed images have been segmented with the use of the algorithm of hierarchical clusterization, presented in the eCognition 3.0 software. For this purpose, all the bands (except thermal) of Landsat TM and ETM+ images, the threshold characteristic, contiguity matrix and textural characteristic “homogeneity” were used. That allowed to classify remotely sensed images with high accuracy and to get time series thematic maps with populated and unpopulated land use types.

Accuracy criteria. A serious obstacle in the dynamics analysis and forecasting of the land cover change is the unresolved problem. Up to now it should be found: what is the best way of model adequacy evaluation and which criteria should be used for modeling (forecast) results of accuracy estimation. These problems are connected with the necessity to obtain true data for the appropriate time moment. However, the quantitative accuracy estimation can not always adequately reflect the processes which take place on the land cover (e.g. Jenerette, 2001). It should be noted that visual and numerical evaluation in some cases can give significantly various results.

A lot of criteria are applied for numerical accuracy estimation for classifying space images and land cover changes modeling. Such criteria as *kappa index of agreement*, *overall accuracy*, *Kno*, *Klocation* and *Khisto* (e.g. Hagen, 2002) are the most used ones by researchers. Different criterion allows taking into account the different features of landscape types. For example, *Klocation* shows land cover cells location features of comparing thematic maps; *Khisto* provides quantitative comparison of land cover types at these maps. Thus, *kappa* is calculated as a product of *Klocation* and *Khisto* allows considering both spatial and quantitative features of comparing maps. In this connection, it is more reasonable to use all mentioned criteria to see different features of comparing land cover maps in order to facilitate the modeling accuracy estimation problem.

Suitability map designing. Available prior information about factors influencing the growth (depression) of the investigated landscape types is essential for the use of adequate land cover change modeling. Formalization of this information is carried out by means of spatial analysis functions within the GIS by probabilistic suitability maps designing (e.g. Zamyatin, 2006). Five factors that influence positively urban growth have been considered to build the suitability map for 2001: distance to main roads, distance to Lisbon city center, distance to centers of Sintra and Cascais, slope and existing urban areas of 1989. These factors were standardised using buffers and regression analysis according to a percentage of urbanized pixels between 1989 and 1994. The map for the main roads and the distance from city center was derived from the *Carta Itinerária Militar de Portugal* of the Portuguese Army Geographical Institute. The slope map was derived from *NASA'Shuttle Radar Topography Mission* data. The slope map was reclassified into 10 percentage slope intervals. Existing urban areas of 1989 received a value of maximum suitability. The five factors were linearly combined, using equal weights, with the suitability map that represents the suitability of each location to be urbanized between 1989 and 2001 (Fig. 2).

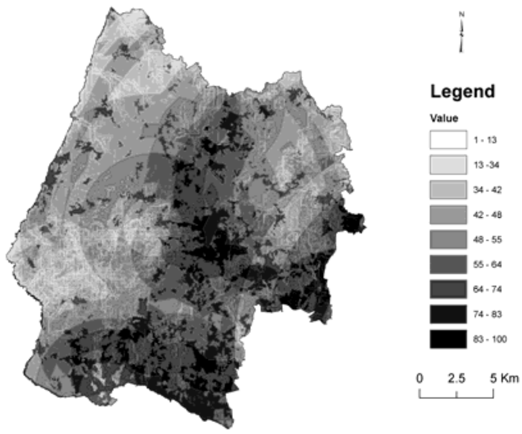


Figure 2: The suitability map with probabilities of changing to urban between 1989 and 2001.

Used models. The application for solving land cover change forecasting problem was made by two mentioned different types of the models, namely: *CA_Markov* realized in GIS *Idrisi Kilimanjaro* and the original model developed by the authors and conditionally named as *CA_Advanced*. The computing part of this model was realized in Microsoft Visual C ++ 6.0 environment with a good binary code optimization and flexible work with memory. Moreover, *CA_Advanced* model allows modeling procedure on the basis of the proposed spatial metrics approach to CA transition rules definition.

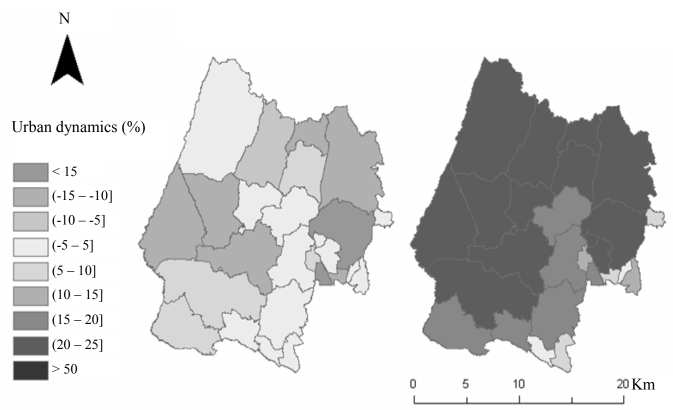
Results. As it was mentioned above, different criteria were used for accuracy estimation of models. The numerical estimation of changes occurred within the period from 1989 to 2001 with the best fit sliding window size can be found in the table:

Models / Accuracy	*CA_Markov* size 5×5	*CA_Advanced* size 3×3	*"Null model"* 1994
Overall accuracy	83,59	**83,95**	80,00
Kappa	64,82	**66,28**	57,35
Kno	67,17	**67,90**	59,99
Klocation	75,27	68,57	64,57
Khisto	86,12	**96,66**	88,82

Results represented in the table show rather high parameters for "*Null model*" column that populated area had not been changed essentially since 1994 to 2001. This confirms the complexity of the model estimation process. Despite the noted complexity, all accuracy estimation results obtained for the model *CA_Advanced* (allocated by bold) are a little bit better than those ones for the two other models. It allows drawing a conclusion about the expediency and efficiency of the proposed approach to CA modeling with the use of spatial characteristics (3)–(6). Except of this, it is possible to note perspectives of carrying out more complex additional research on the self-descriptiveness analysis of spatial characteristics for CA transition rule definition.

Long-term forecasting. Tasks of the land cover dynamics analysis assume carrying out both short-term (2-5 years) and long-term (5 and more years) forecasts. It should be noted that practical model application which is suitable for the short-term forecast, may be not suitable for long-term and vice-versa. Despite the essential unpredictability of land cover development in the long-term period, the long-term forecast with the use of the efficient model can be useful and will allow an expert to reveal and estimate existing land cover tendencies and take steps to decrease consequences of possible crisis situations.

The research provided a number of experiments for the long-term forecasting. One of the examples of research conducted in 2025 on models *CA_Markov* and *CA_Advanced* is presented in Fig. 3. For the correct comparative analysis of modeling results, the equal initial data for both models were used. Available raster maps of the populated territories and the cadastral information made possible to present researched area in the form of 26 administrative and territorial units with the estimated changes in each of them.



a *b*

Figure 3: The long-term forecasting in 2025 for two models: *a) CA_Markov, b) CA_Advanced*

The obtained results help to reveal the insufficient *CA_Markov* model effectiveness for the long-term forecasting (Fig. 3, *a*), because the modeling process does not demonstrate consecutive urban growth process as in reality. At the same time, *CA_Advanced* application for the long-term forecasting has more realistic and appropriate results (Fig. 3, *b*).

CONCLUSION

Even today land cover change analysis problems, short-term and long-term forecasting of the urban growth process are extremely complex. Existing algorithms and software do not always allow solve efficiently such problems. The paper is oriented towards the search for more effective modeling and the land cover change forecasting. For land cover change forecasting the spatial metrics approach to cellular automata transition rules was proposed and implemented. This approach allows a more effective consideration of the land spatial characteristics and the increase of adequacy of final cellular automata modeling. The comparative analysis of models *CA_Markov* and *CA_Advanced* is carried out for solving land cover change analysis problems, short-term and long-term forecasting of urban growth areas in Sintra-Cascais (Portugal). It shows perspectives of the proposed approach, especially for the long-term forecast problem.

The authors would like to acknowledge to Professor N. Markov for helpful discussion and support of the research. Supported by the Russian Foundation for Basic Research, grant number 06-07-89190.

BIBLIOGRAPHY

Briassoulis H. Analysis of Land Use Change: Theoretical and Modeling Approaches. The Web Book of Regional Science, Regional Research Institute, West Virginia University, 1999, http://www.rri.wvu.edu /WebBook/ Briassoulis/ contents.htm (01.07.2006).

Baker W.L. A review of models of landscape change // Landscape Ecology, 1989, № 2, P. 111–133.

Cabral P., Geroyannis H., Gilg J, Painho M. Analysis and modeling of land-use and land-cover change in Sintra-Cascais area // 8th AGILE Conference on Geographic Information Science, 2005, P. 51–60.

Clark Labs. GIS Idrisi 32. – http://www.clarklabs.org/IdrisiSoftware/ (01.07.2006).

Fulong W. Calibration of stochastic cellular automata: the application to rural-urban land conversions // int. j. Geographical Information Science, 2002, vol. 16, no. 8. P. 795–818.

Hagen A. Multi-method assessment of map similarity // 5th AGILE Conference on Geographic Information Science, 2002, P. 1–8.

Jenerette G.D., J. Wu. Analysis and Simulation of Land Use Change in the Central Arizona – Phoenix region // Landscape Ecology, 2001, V. 16. – P. 611–626.

O'Sullivan D., Torrens P. M. Cellular models of urban systems // Theoretical and Practical Issues on Cellular Automata. London: Springer-Verlag, 2000, P. 108–117.

Pontius Jr., J. Malanson. Comparison of the structure and accuracy of two land change models // International Journal of Geographical Information Science, 2005, V. 19. – № 2. – P. 243–265.

Verburg P.H., e.a., A method to analyse neighborhood characteristics of land use patterns // Computers, Environment and Urban Systems, 2003, №3. – P.354–369.

Zamyatin A., Mikhaylov P. Technology of suitability maps designing for land use and land cover change modelling – http://www.agile2006.hu/papers_link.htm // 9 AGILE Conference on Geographic Information Science, 2006. (01.07.2006).

Zamyatin A., Markov N. Approach to Land Cover Change Modelling Using Cellular Automata // Proceedings of 8th Conference on Geographic Information Science. – Estoril.– Portugal.– AGILE, 2005, P.587-592.