Anais XI SBSR, Belo Horizonte, Brasil, 05 - 10 abril 2003, INPE, p. 1687 - 1694.

SIMULATING URBAN LAND USE CHANGE THROUGH CA-BASED MODELING AND LOGISTIC REGRESSION

CLÁUDIA MARIA DE ALMEIDA [1]
ANTONIO MIGUEL VIEIRA MONTEIRO [1]
GILBERTO CÂMARA [1]
BRITALDO SILVEIRA SOARES-FILHO [2]
GUSTAVO COUTINHO CERQUEIRA [2]
WILLIAM LEITE ARAÚJO [2]
ALEXANDRE EDUARDO PANTUZZO [3]

[1]INPE - Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brazil
almeida@ltid.inpe.br, {miguel, gilberto}@dpi.inpe.br

[2]Centro de Sensoriamento Remoto - Universidade Federal de Minas Gerais
Av. Antonio Carlos, 6627 - 31270-900 – Belo Horizonte - MG, Brazil
{britaldo, cerca, araujo}@csr.ufmg.br

[3]Consult Brasil – Tecnologia e Negócios
Av. do Contorno, 2090 – 9° andar - 30110-070 – Belo Horizonte - MG, Brazil
pantuzzo@consultbrasil.com.br

Abstract. Simulating urban land use change through stochastic methods invariably demands the assessment of spatial land use transition probabilities. This has been accomplished to date mostly by empirical calculations and statistical linear methods. In the present work, we introduce a framework for simulating urban land use dynamics based on the estimation of land use transition probabilities through logistic regression. These probabilities drive a cellular automaton (CA) simulation model, based on eight cell Moore neighborhoods and stochastic transition algorithms. A medium-sized town in the west of São Paulo State, *Bauru*, was adopted as case study. Different simulation outputs for the case study town in the period 1979-1988 were generated, and statistical validation tests were then conducted for the best results, employing a multiple resolution fitting procedure.

Keywords: urban modeling, land use dynamics, logistic regression, cellular automata, town planning.

1. Introduction

Cellular automata (CA) models have found applications in diverse fields, ranging from statistical and theoretical physics to land use and land cover change, traffic engineering and control, diseases spread, behavioral biology, amongst others. The basic idea of these models is very simple: in a gridded space (raster) a series of transition rules are enforced to govern the state of a randomly placed cell depending on the configuration of its neighborhood. The simplicity, operational and mathematical tractability of these models as well as their surprising capacity to reproduce complexity embedded in processes of spatial change as reflected in emergent phenomena account for their wide popularity and use in the applied sciences.

More recently, cellular automata have found their way into 2-D applications in urban modeling (Batty, 2000). There are currently some twenty or more applications of CA to cities such as the diffusion or migration of resident populations (Portugali et al., 1997) the competitive location of economic activities (Benati, 1997), the joint expansion of urban surface and traffic network (Batty and Xie, 1997), generic urban growth (Clarke et al., 1997), urban land use dynamics (White and Engelen, 1997), and so forth.

Specifically regarding urban land use dynamics, CA models can be basically subdivided into either predominantly deterministic or stochastic. A good representative of predominantly deterministic CA models is the urban growth study for the San Francisco Bay area, conducted by Clarke et al. (1997). Although this model incorporates a certain randomness in selecting the cells for urban growth and in promoting the spread of growth seeds, its transition rules are fundamentally deterministic in the sense that the cell suitability for being urbanized is not dependent upon probabilistic methods.

An illustrative example of the second category of models is the SIMLUCIA, conceived by White et al. (1997). The stochastic feature of this model is present in the calculation of land use transition probabilities for each cell, which is basically a function of the cell suitability for the new activity in question and its relative accessibility for such an activity. The formula adopted for the probabilities estimation is a combined linear product of factors affecting land use change plus a disturbance term.

Logistic regression has been used for simulating urban land use change in a few cases. Morisette et al. (1999) and Jianquan and Masser (2002) conduct modelings of urban development patterns, which are not CA-based and deal with only two categories of land use (urban and non-urban). Tang and Choy (2000) and Wu (2000) applied the logistic regression method to more specific land use transition issues, such as office development and industrial firm location respectively, but their experiments are not carried out in CA environments either.

This paper addresses the simulation of land use change for sub-categories of urban land use (e.g. residential, commercial, industrial, etc.) by means of logistic regression and CA-based modeling. The following section introduces the study area and pre-processing techniques. The third section presents exploratory analysis procedures and approaches the intervening factors in urban land use change. In the forth section, the theoretical structure of the model is presented, providing an overview on the logistic regression method and explaining how it is used for determining transition probabilities governing changes in land use as functions of a variety of socio-economic and infrastructural factors. Section five discusses questions related to model implementation and calibration. In section six, the simulation results are presented and explained in the context of urban land use dynamics. Finally, section seven is reserved for final comments and directions for future work.

2. The Study Area

This simulation model is developed for the city of Bauru, which is located in West São Paulo State, and in 2000 had a rapidly growing population of 309,640. The period for which the model is fitted is from 1979 to 1988 when the population grew from 179,823 to 232,005.

The city maps provided by the Bauru local authorities presented inconsistencies due to the fact that illegal settlements are not shown on the official maps, and not all of the legally approved settlements drawn have been in fact implemented. Moreover, some urban zones refer to areas which are not yet occupied, and some other zones categories do not correspond to the prevailing use indeed encountered within their limits, reflecting just the local officials´ intention for their future use. In this way, satellite imagery arise as a feasible solution for the identification of urban settlements actually existent, as well as for the delineation of the true urban occupation boundaries of the case study town.

In this way, the initial (1979) and final (1988) land use maps were subjected to a reclassification of zones according to their dominant effective use; residential zones of different densities were all reclassified to simply residential, and special use and social infrastructure were reclassified to institutional. Eight land use zone categories were thence defined, namely: residential, commercial, industrial, services, institutional, mixed use, leisure/recreation, and the all-embracing non-urban land use. Districts segregated from the main urban agglomeration by more than 10 km were judged outside the simulation area, and the traffic network was not considered to be at a fine enough scale to be represented as a land use.

The land use maps for the two time slices are shown in **Figure 1 (a)** and **(b)**. The changes between 1979 and 1988 are shown in **Figure 2 (a)** with the most significant land use change – from non-urban to residential use – shown in **Figure 2 (b)**. All data used in this experiment were represented at 100 m x 100 m grid square, pre-processed using the SPRING GIS (from the Division for Image Processing of the Brazilian National Institute for Space Research – DPI-INPE) and IDRISI (from Clark University).

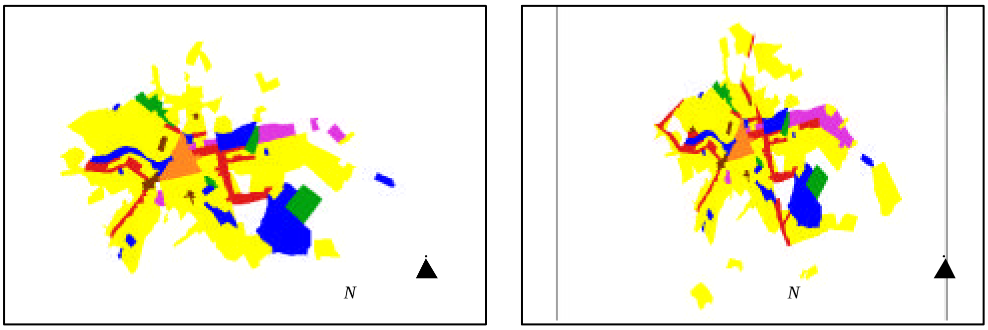


Figure 1. (a) Land use in Bauru in 1979 (left) and (b) land use in Bauru in 1988 (right). Residential use is yellow, commercial use is orange, institutional use is blue, industrial use is purple, services corridors are red, leisure/recreation is green, mixed use is brown, and non-urban use is white.

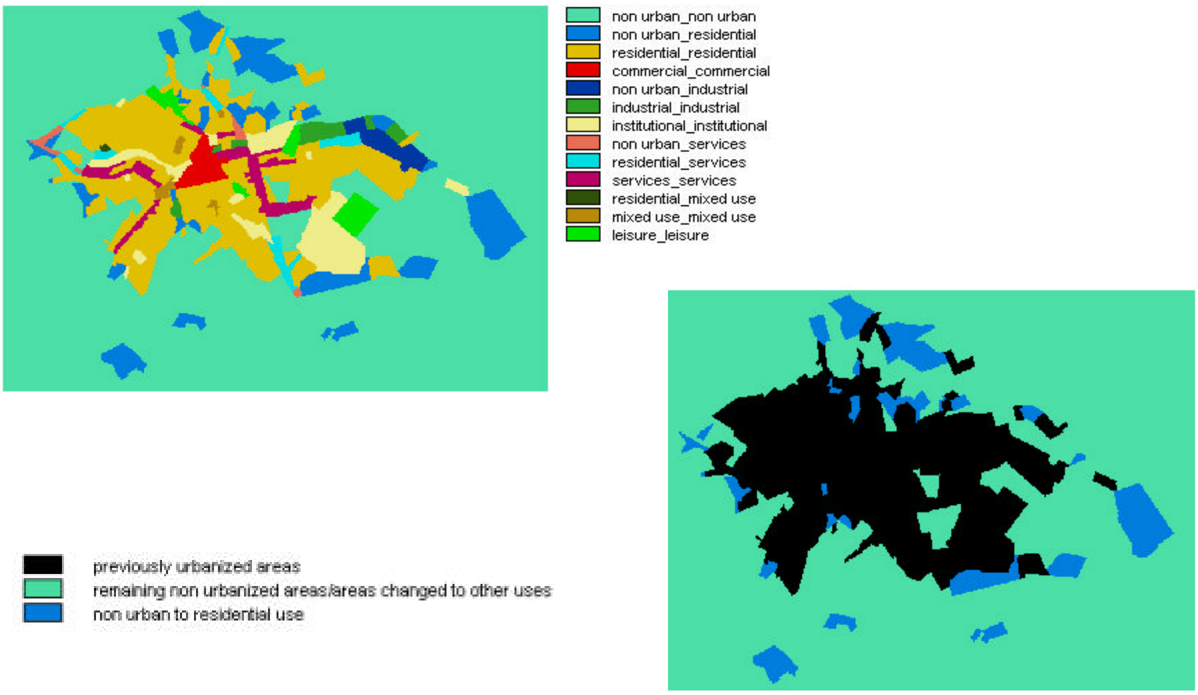


Figure 2. (a) Land use change from 1979 to 1988 (top left) and (b) Map showing the transition non-urban to residential land use during the period 1979-1988 (bottom right).

3. Exploratory Analysis and Selection of Variables

From the map of land use changes from 1979 to 1988, shown in Figure 2 (a), obtained through a cross-tabulation operation between the initial and final land use maps shown in Figure 1, five types of transitions were observed and are listed in **Table 1**.

Table 1
Observed land use transitions

Notation	Land Use Transition
NU_RES	Non-Urban to Residential
NU_IND	Non-Urban to Industrial
NU_SERV	Non-Urban to Services
RES_SERV	Residential to Services
RES_MIX	Residential to Mixed Use

To explain each of the five existent land use transitions, twelve variables were selected from an initial bunch of over forty variables regarding infrastructural and socio-economic aspects of Bauru. Examples of maps of independent variables are shown in **Figures 3 (a)** and **(b)**.

Figure 3. (a) Existence of social housing units and (b) Map of distances to industrial zones (right).

The map shown in Figure 3 (a) is a typical example of a discrete or categorical variable, which assumes values 1 for those cells where social housing is found and 0 otherwise. The map of distances in Figure 3 (b) can be either treated as a continuous variable (real values grid) or a categorical one (ranges of distances). In the particular case of this experiment, all variables related to maps of distances were treated as categorical, since practically all of them presented a non-linear and/or multimodal behavior in relation to the respective land use transitions. Handling this information as a continuous variable implies a great data heterogeneity, what would certainly bring about noise in the simulations and harm the model calibration.

Empirical procedures were used for variables selection, like the visualization of distinct variables superimposed on the final land use map, what aimed at identifying the set of those ones more meaningful to explain the five different types of land use change. Another auxiliary method was the analysis of boxplots generated by each selected independent variable versus the respective land use transition. **Figures 4 (a)** and **(b)** shows examples of boxplots for the variables presented in Figure 3 (a) and (b). It is observable for the first case that the majority of the cells where the transition from residential to mixed use occurs coincide with the cells where social housing units are also found. In the second case, the boxplot shows that the transition from non-urban to industrial use takes place in the closest areas from the already existent industrial zones. Both types of analyses (visualization of maps overlay and boxplots) led to a preliminary selection of independent variables, as shown in **Table 2**.

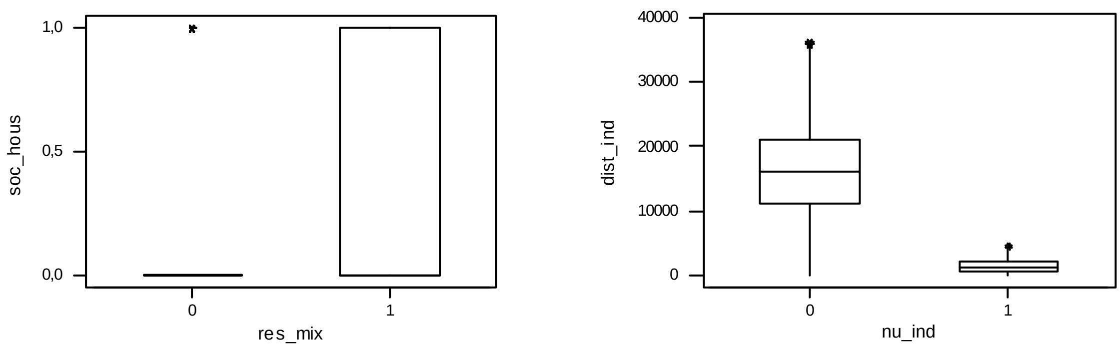


Figure 4. (a) Boxplot of the transition residential to mixed use versus social housing (left) and (b) Boxplot of the transition non-urban to industrial use versus distance to industrial zones (right).

Table 2
Selection of variables determining land use change

Independent Variables (Notation)	NU_RES	NU_IND	NU_SERV	RES_SERV	RES_MIST
Area served by water supply (water)			♦		
Medium-high density of occupation: 25% to 40% (mh_dens)					♦
Existence of social housing (soc_hous)					♦
Distances to ranges of commercial concentration (com_kern)	♦		♦		
Distances to industrial zones (dist_ind)		♦			
Distances to residential zones (dist_res)			♦		
Distances to peripheral residential settlements (per_res)	♦				
Distances to isolated institutional use (dist_inst)	♦				
Distances to main existent roads (exist_rds)	♦				
Distances to the service and industrial axes (serv_axes)		♦	♦	♦	
Distances to planned roads (plan_rds)					♦
Distances to peripheral roads (per_rds)	♦				♦

Have the variables been selected, it becomes then necessary to check for their spatial dependence or association. This is done for all possible pairwise combination of variables existent in each of the five land use transitions separately. For this end, the Cramer´s statistic (V) and the Joint Information Uncertainty (U) indices (Bonham-Carter, 1994) were used. Values less than 0.5 suggest less association rather than more. As none of the association values surpassed this threshold, no variables initially selected for the modeling experiment have been discarded from the analysis.

4. Methods: A Logistic Regression-Based Cellular Automaton Model

4.1 Global Transition Dynamics

Probabilities of land use transition were initially calculated for the whole study area in absolute terms, i.e. without the influence of socio-economic or infrastructural factors. This has been accomplished through a cross-tabulation operation between the initial and final land use maps shown in Figure 1. The probabilities estimates are presented in **Table 3**.

Table 3
Global transition probabilities for Bauru, 1979-1988

Land use	Non-Urban	Residential	Commercial	Industrial	Institut	Services	Mixed	Leis/Rec
Non-Urban	0.9171331	0.0697519	0	0.0095301	0	0.0035848	0	0
Residential	0	0.9379833	0	0	0	0.0597520	0.0022647	0
Commercial	0	0	1	0	0	0	0	0
Industrial	0	0	0	1	0	0	0	0
Institutional	0	0	0	0	1	0	0	0
Services	0	0	0	0	0	1	0	0
Mixed	0	0	0	0	0	0	1	0
Leisure/Rec	0	0	0	0	0	0	0	1

4.2 Local Transition Dynamics

A customized reckoning of transition probabilities was then conducted at the cellular level, taking into account the local socio-economic and infrastructural variables. Each land use transition was separately modeled in these statistical calculations, what complies with the algorithmic logic of the modeling software, in which each transition has its calibration parameters individually adjusted.

The binary logistic regression model has been adopted. Each transition is coded as 1 and permanence in the original state as well as changes to uses other than the one considered in the transition were coded as 0. Thus, a change in the cell land use during the simulation period is dependent on its initial state as well as on its P_{ij} *(x,y)*, which is the probability that a cell at position *(x,y)* will change from state *i* to state *j*. The dependence of the local transition probabilities P_{ij} *(x,y)* on each independent variable V_n *(x,y)* is estimated by the logistic model:

$$L = log \left(\frac{P_{ij}(x,y)}{(1 - P_{ij}(x,y))} \right) = ß_{0,ij} + ß_{1,ij} V_{1,xy} + \ldots + ß_{k,ij} V_{k,xy} \qquad (1)$$

what implies that

$$P_{ij}(x,y) = \frac{e^{L}}{1 + e^{L}} \qquad (2)$$

The logistic regression models for each of the five transitions included the sets of variables as shown in Table 2 and excluded the least significant variable (if any) at each step. Significance was based on the Wald chi-square test and the G statistic. The model is accepted when all independent variables are significant at the 0.05 level and the loss of the G statistic remains lower than 5%. The parameters values for each transition are shown in **Table 4** and were obtained through the maximum likelihood method using the statistical package MINITAB, release 13.0. Although the variables "*dist_res*" and "*mh_dens*" were not significant at the 0.05 level, they were kept in the model in view of their effective contribution for explaining the transitions "*nu_serv*" and "*res_mix*", respectively. According to Hosmer and Lemeshow (1989), "we must not base our models entirely on tests of statistical significance, since there are numerous other considerations that will influence our decision to include or exclude variables from a model".

Table 4
Results of the logistic regression analyses for Bauru, 1979-1988

VARIABLES	*Transition NU_RES*		*Transition NU_IND*		*Transition NU_SERV*		*Transition RES_SERV*		*Transition RES_MIX*	
	ßk	*P*	*ßk*	*P*	*ßk*	*P*	*ßk*	*P*	*ßk*	*P*
Constant ($ß_0$)	7.646900	0.000	5.274530	0.000	4.865300	0.000	-1.551900	0.000	3.901200	0.000
water	#	#	#	#	#	#	1.708810	0.000	#	#
mh_dens	#	#	#	#	#	#	#	#	0.383300	0.232
soc_hous	#	#	#	#	#	#	#	#	-1.068800	0.000
com_kern	-0.924990	0.000	#	#	-1.461660	0.000	#	#	#	#
dist_ind	#	#	-1.048320	0.000	#	#	#	#	#	#
dist_res	#	#	#	#	0.027680	0.442	#	#	#	#
per_res	-0.392090	0.000	#	#	#	#	#	#	#	#
dist_inst	-0.405525	0.000	#	#	#	#	#	#	#	#
exist_rds	0.051476	0.000	#	#	#	#	#	#	#	#
serv_axes	#	#	-0.741110	0.000	-0.974470	0.000	-0.929550	0.000	#	#
plan_rds	#	#	#	#	#	#	#	#	-1.865200	0.000
per_rds	-0.309469	0.000	#	#	#	#	#	#	-0.521040	0.000
					RESULTS FOR THE TESTS OF GOODNESS – OF - FIT					
Tests	*Chi-square*	*P*	*Chi-square*	*P*	*Chi-square*	*P*	*Chi-square*	*P*	*Chi-square*	*P*
Pearson	41,202.475	0.000	13,639.316	0.000	938.120	0.000	338.064	0.000	422.206	0.000
Deviance	30,435.653	0.000	6,055.790	0.000	774.369	0.000	341.693	0.000	328.558	0.000
Hosmer-Lem.	613.082	0.000	258.618	0.000	44.667	0.000	247.916	0.000	1.653	0.438

5. Model Calibration

By means of the parameters estimated in the logistic regression analyses, the simulation model – DINAMICA – developed at the Center for Remote Sensing of the Federal University of Minas Gerais (CSR-UFMG), will calculate the cells transition probabilities and generate maps of probabilities for each of the five types of land use change. These maps are compared to the actual land use transitions (**Figure 5**), and both of them together with preliminary simulation results are used for the model calibration.

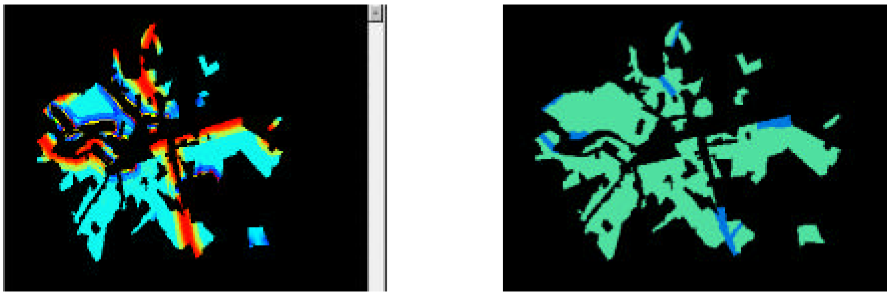
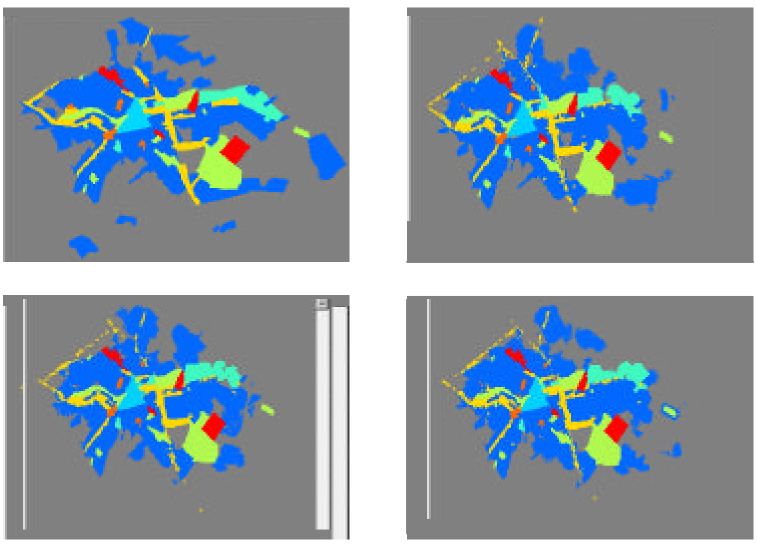


Figure 5. (a) Map of transition probabilities: res_serv (left) and (b) Map of land use transition: res_serv (right).

The calibration process not only defines the best set of variables to explain each of the transitions but also internal parameters of the DINAMICA model like number of iterations, average size and variance of patches, etc. Have a good calibration been achieved, DINAMICA will carry out the final runs, where changes in the cells states occur through two types of transition algorithms based on eight cell Moore neighborhoods and which employ a stochastic selecting mechanism: (i) the "*expander*", which accomplishes transitions from a state *i* to a state *j* only in the adjacent vicinities of cells with state *j* ; and (ii) the "*patcher*", which realizes transitions from a state *i* to a state *j* only in the adjacent vicinities of cells with state other than *j* (Soares-Filho et al., 2002).

6. Results and Discussion

The best three simulation results are shown in **Figure 6** and are seen in ERMapper, employed by DINAMICA as a visualization device.



(a) *(b)*

(c) *(d)*

Figure 6. (a) Real land use change in Bauru from 1979 to 1988 (top left) and Simulations (b) 1, (c) 2 and (d) 3.

These three simulations were validated according to a multiple resolution fitting method (Constanza, 1989), and the values for goodness of fit obtained for windows sizes of 3x3, 5x5 and 9x9 cells were 0.905172, 0.907539, and 0.907868, respectively for S1, S2, and S3. It is observable that the land use transitions comply with economic theories of urban growth and change, where there is a continuous search for optimal location, able to assure competitive real state prices, good accessibility conditions, rationalization of transportation costs, and a strategic location in relation to suppliers and consumers markets (Almeida et al., in press).

6. Conclusions

Methods of open systems modeling of which CA is one of the best examples and which meet many requirements for simulating dynamic processes quickly and efficiently are rarely implemented in GIS. As a result, GIS remains surprisingly narrowly focused (Openshaw, 2000). In this way, our group at DPI-INPE is currently committed to the development of a multi-purpose 2D and 3D land use CA simulation module to be integrated with SPRING GIS.

References

Almeida, C. M.; Batty, M.; Monteiro, A. M. V.; Câmara, G.; Soares-Filho, B. S.; Cerqueira, G. C.; Pennachin, C. L. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation, *Computers, Environment and Urban Systems*, in press.

Batty, M. Geocomputation using cellular automata. In: Openshaw, S.; Abrahart, R. J. (eds), *Geocomputation*, New York: Taylor & Francis, chap.5, p. 95-126, 2000.

Batty, M.; Xie Y. Possible urban automata, *Environmental and Planning B*, v.24, p. 175-192, 1997.

Benati, S. A cellular automaton for the simulation of competitive location, *Environmental and Planning B*, v.24, p. 205-218, 1997.

Bonham-Carter, G. F. *Geographic information systems for geoscientists: modelling with GIS*, New York: Pergamon, 1994.

Clarke, K. C.; Hoppen, S.; Gaydos, L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area, *Environmental and Planning B*, v.24, p. 247-261, 1997.

Constanza, R. Model goodness of fit: a multiple resolution procedure, *Ecological Modelling*, v.47, p. 199-215, 1989.

Jianquan, C.; Masser, I. *Towards a Spatial Analysis Framework: Modelling Urban Development Patterns*, [online], <http://www.geocomputation.org/2001/papers/cheng.pdf>, Oct. 2002.

Hosmer, D. W.; Lemeshow, S. *Applied Logistic Regression*, New York: John Wiley & Sons, 307p., 1989.

Morisette, J. T.; Khorram, S.; Mace. T. Land-cover change detection enhanced with generalized linear models, *International Journal of Remote Sensing*, v.20, p.2703-2721, 1999.

Openshaw, S. GeoComputation. In: Openshaw, S.; Abrahart, J. (eds), *Geocomputation*, London: Taylor & Francis, p. 1-32, 2000.

Portugali, J.; Benenson, I.; Omer, I. Spatial cognitive dissonance and sociospatial emergence in a self-organizing city, *Environmental and Planning B*, v.24, p. 263-285, 1997.

Soares-Filho, B. S., Cerqueira, G. C.; Pennachin, C. L. DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier, *Ecological Modelling*, v.154, p.217-235, 2002.

Tang, B. S.; Choy, L. H.-T. Modelling planning control decisions: a logistic regression analysis on office development applications in urban Kowloon, Hong Kong, *Cities*, v.17, p.219-225, 2000.

White, R. W.; Engelen, G. Cellular automaton as the basis of integrated dynamic regional modelling, *Environmental and Planning B*, v.24, p. 235-246, 1997.

Wu, F. Modelling intrametropolitan location of foreign investment firms in a Chinese city, *Urban Studies*, v. 37, p. 2441-2464, 2000.