Modeling the Urban Evolution of Land Use Transitions Using Cellular Automata and Logistic Regression

C. M. Almeida, A. M. V. Monteiro, G. Câmara,
Division for Image Processing - DPI
Brazilian National Institute for Space Research - INPE
São José dos Campos – SP, Brazil
almeida@ltid.inpe.br, miguel@dpi.inpe.br,
gilberto@dpi.inpe.br

B. S. Soares-Filho, G. C. Cerqueira,
W. L. Araújo, A. E. Pantuzzo
Center for Remote Sensing - CSR
Federal University of Minas Gerais – UFMG
Belo Horizonte – MG, Brazil
britaldo@csr.ufmg.br, cerca@csr.ufmg.br,
araujo@csr.ufmg.br, pantuzzo@consultbrasil.com.br

Abstract—**The present work introduces a framework for simulating urban land use dynamics based on the estimation of land use transition probabilities through logistic regression. Different simulation outputs for a case study town, Bauru, 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; celllular automata; town planning

I. 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. More recently, cellular automata have found their way into 2-D applications in urban modeling [1], and there are currently some twenty or more applications of CA to cities, like traffic simulation, segregation, gentrification, land use dynamics, etc.

Specifically regarding urban land use dynamics, CA models can be basically subdivided into either deterministic or stochastic. Simulating urban land use change through stochastic methods invariably demands the assessment of spatial land use transition probabilities. Logistic regression has been used to model urban land use change in a few cases. References [2] and [3] conduct simulations of urban development patterns, which are not CA-based and deal with only two categories of land use (urban and non-urban). Reference [4] and Reference [5] 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.

II. METHODS

A. Exploratory Analysis and Selection of Variables

From a map of land use changes from 1979 to 1988, obtained through a cross-tabulation operation between the initial and final land use maps, five types of transitions were observed: (i) non-urban to residential use (nu_res); (ii) non-urban to industrial use (nu_ind); (iii) non-urban to services (nu_serv); (iv) residential to services (res_serv), and (v) residential to mixed use (res_mix).

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.

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 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 (Fig. 1).

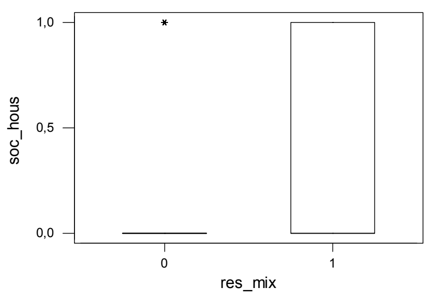

Figure 1. Boxplot of the transition “res_mix” versus social housing.

Both types of analyses (visualization of maps overlay and boxplots) led to a preliminary selection of independent variables. 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, the Joint Information Uncertainty [6] as well as the Correlation Indices were used. Values less than 0.5 suggest less association rather than more. Since none of the values surpassed this threshold simultaneously for the three indices considered, no variables initially selected for modeling have been discarded from the analysis.

B. 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 (1979) and final (1988) land use maps.

C. 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 - DINAMICA, 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:

, (1)

what implies that:

. (2)

The logistic regression models for each of the five transitions included the preliminarily selected sets of variables 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 for each transition and their respective Wald test p-value are shown in Table 1 and were obtained through the maximum likelihood method using the statistical package MINITAB, release 13.0. Although the variables “distances to residences” and “medium-high density of occupation” 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 [7], “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.”

D. Model Calibration

By means of the parameters estimated in the logistic regression analyses, the simulation model – DINAMICA – will calculate the cells transition probabilities and generate maps of probabilities (Fig. 2) for each of the five types of land use change. These maps are compared to the actual land use transitions (Fig. 3), and both of them together with preliminary simulation results are used for the model calibration.

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* [8].

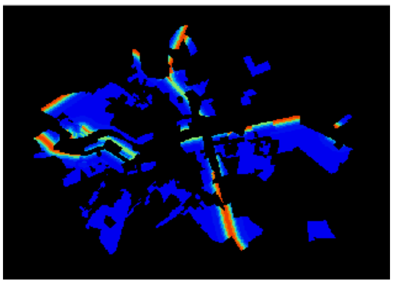

Figure 2. Example of a map of probabilities for the transition “res_serv”. High probabilities are in mid gray.

Figure 3. Example of a land use transition map: “res_serv”. Areas of transition are in light gray.

TABLE 1
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*	
	βk	*P*	*βk*	*P*	*βk*	*P*	*βk*	*P*	*βk*	*P*
Constant (β_0)	7.646900	0.000	5.274530	0.000	4.865300	0.000	-1.551900	0.000	3.901200	0.000
Water supply	#	#	#	#	#	#	1.708810	0.000	#	#
Medium-high density	#	#	#	#	#	#	#	#	0.383300	0.232
Social housing	#	#	#	#	#	#	#	#	-1.068800	0.000
Distances to commerce	-0.924990	0.000	#	#	-1.461660	0.000	#	#	#	#
Distances to Industries	#	#	-1.048320	0.000	#	#	#	#	#	#
Distances to Residences	#	#	#	#	0.027680	0.442	#	#	#	#
Dist. to Per. Settlements	-0.392090	0.000	#	#	#	#	#	#	#	#
Dist. to Institutional Use	-0.405525	0.000	#	#	#	#	#	#	#	#
Distances to Exis. Roads	0.051476	0.000	#	#	#	#	#	#	#	#
Dist. to Services Axes	#	#	-0.741110	0.000	-0.974470	0.000	-0.929550	0.000	#	#
Dist. to Planned Roads	#	#	#	#	#	#	#	#	-1.865200	0.000
Dist.to Peripheral Roads	-0.309469	0.000	#	#	#	#	#	#	-0.521040	0.000

III. RESULTS AND DISCUSSION

The simulation outputs are seen in ERMAPPER, employed by the DINAMICA model as a visualization device. Even though the internal parameters of the model are kept unchanged, different runs will produce different results, given the stochastic nature of this simulation software. Fig. 4 and 5 respectively present the real and a simulated land use map for Bauru in 1988.

This simulation result was validated according to a multiple resolution fitting method [9], and the value for goodness of fit obtained for window sizes of 3x3, 5x5 and 9x9 cells was 0.907868. 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.

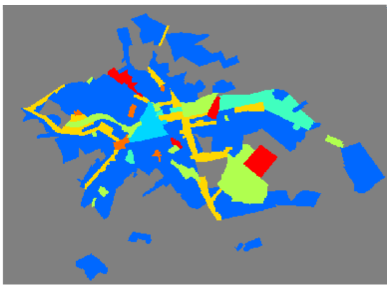

Figure 4. Real land use map for the case study town of Bauru in 1988.

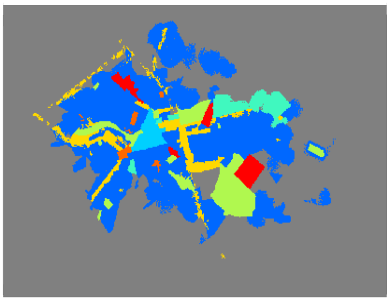

Figure 5. Simulated land use map for the case study town of Bauru in 1988.

IV. 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 [10]. As a result, GIS remains surprisingly narrowly focused [11]. In this way, our group at DPI-INPE is currently committed to the development of a flexible multi-scale and multi-purpose 2D and 3D CA simulation module to be available at "Terralib", an open source interoperable library.

To finalize, it is worth stressing here the wide feasibility to optimize the simulations by means of a model which embraces more refined algorithmic logics (fractal parameters, semi-stochastic rules, etc.), suitable for the urban phenomena modeling under consideration.

REFERENCES

[1] M.Batty, "Geocomputation using cellular automata," in *Geocomputation*, S. Openshaw and R. J. Abrahart, Eds. New York: Taylor & Francis, 2000, pp. 95-126.

[2] J. T. Morisette, S. Khorram, and T. Mace, "Land-cover change detection enhanced with generalized linear models," *International Journal of Remote Sensing*, vol. 20, pp. 2703-2721, 1999.

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

[4] B. S. Tang and L. H.-T. Choy, "Modelling planning control decisions: a logistic regression analysis on office development applications in urban Kowloon, Hong Kong," *Cities*, vol. 17, pp. 219-225, 2000.

[5] F. Wu, "Modelling intrametropolitan location of foreign investment firms in a Chinese city," *Urban Studies*, vol. 37, pp. 2441-2464, 2000.

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

[7] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. New York: John Wiley & Sons, 1989.

[8] B. S. Soares-Filho, G. C. Cerqueira, and C. L. Pennachin, "DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier," *Ecological Modelling*, vol. 154, pp. 217-235, 2002.

[9] R. Constanza, Model goodness of fit: a multiple resolution procedure," *Ecological Modelling*, vol. 47, pp. 199-215, 1989.

[10] P. A. Burrough, "Dynamic Modelling and Geocomputation," in *Geocomputation: A Primer*, P. A. Longley, S. M. Brooks, R. McDonnell, and B. MacMillan, Eds. Chichester, UK: John Wiley & Sons, 1998, pp. 165-192.

[11] S. Openshaw, "GeoComputation," in *Geocomputation*, S. Openshaw and J. Abrahart, Eds. London: Taylor & Francis, 2000, pp. 1-32.