International Journal of Remote Sensing
Vol. 26, No. 4, 20 February 2005, 759–774

Taylor & Francis
Taylor & Francis Group

GIS and remote sensing as tools for the simulation of urban land-use change

CLÁUDIA MARIA DE ALMEIDA*†, ANTONIO MIGUEL VIEIRA MONTEIRO†, GILBERTO CÂMARA, BRITALDO SILVEIRA SOARES-FILHO‡, GUSTAVO COUTINHO CERQUEIRA‡, CÁSSIO LOPES PENNACHIN§ and MICHAEL BATTY¶

†National Institute for Space Research (INPE), Avenida dos Astronautas, 1758 – 12227-010, São José dos Campos, São Paulo, Brazil

‡Department of Cartography, Federal University of Minas Gerais (UFMG), Avenida Antônio Carlos, 6627 – 31270-900, Belo Horizonte, Minas Gerais, Brazil

§Intelligenesis do Brasil Ltda, Avenida Brasil, 1438-150530140-003, Belo Horizonte, Minas Gerais, Brazil

¶Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London WC1E 6BT, UK

This paper is concerned with building up methodological guidelines for modelling urban land-use change through Geographical Information Systems, remote sensing imagery and Bayesian probabilistic methods. A medium-sized town in the west of São Paulo State, *Bauru*, was adopted as a case study. Its urban structure was converted into a 100 m × 100 m resolution grid and transition probabilities were calculated for each grid cell by means of the 'weights of evidence' statistical method and upon the basis of the information related to the technical infrastructure and socio-economic aspects of the town. The probabilities obtained from there fed a cellular automaton simulation model—DINAMICA—developed by the Centre for Remote Sensing of the Federal University of Minas Gerais (CSR-UFMG), based on stochastic transition algorithms. 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.

This modelling experiment revealed the plausibility of adopting Bayesian empirical methods based on the available knowledge of technical infrastructure and socio-economic status to simulate urban land-use change. It indicates their possible further applicability for generating forecasts of growth trends both for Brazilian cities and cities world-wide.

1. Introduction

Recent generation models of urban dynamics have dealt with diverse themes. According to Batty (2000), there are currently some twenty or more applications of cellular automaton (CA) models to cities, such as in the diffusion or migration of resident populations (Portugali *et al.* 1997), competitive location of economic activities (Benati 1997), joint expansion of urban surface and traffic network (Batty and Xie 1997), generic urban growth (Clarke *et al.* 1997) and urban land-use

*Corresponding author. Email: almeida@ltid.inpe.br

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2005 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160512331316865

dynamics (Deadman *et al*. 1993, Batty and Xie 1994, Phipps and Langlois 1997, White and Engelen 1997, White *et al*. 1998), amongst others.

Specifically regarding urban land-use dynamics, it is possible to identify three basic trends of CA models in respect to their balance between stochasticity and determinism. The first one concerns the predominantly deterministic models, whose most evident representative 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, which can be spontaneous, diffusive, organic or road-influenced, are fundamentally deterministic in the sense that the cell suitability for being urbanized is not dependent upon probabilistic methods.

A second trend relates to the stochastic models with both deterministic estimations of area for land-use transition and deterministic transition algorithms. A good example of this category of models is the SIMLUCIA, conceived by White *et al*. (1998), which is an integrated model of natural and human systems operating at several spatial scales, and was aimed at providing the officials of the Caribbean Island of Santa Lucia with a tool to explore possible environmental, social and economic consequences of hypothesized climate change.

In this model, a sophisticated set of equations that take into account aspects of the natural environment is formulated in order to estimate the impact of economic and demographic changes on land use. The stochasticity of this model is present in the calculation of the probabilities of land-use transition 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. In the SIMLUCIA transition algorithm, cells are ranked by their highest potential and cell transitions begin with the highest-ranked cell and proceed deterministically downwards, until the number of cells demanded by the above-mentioned equations is reached.

A third trend concerns the stochastic CA models, with both stochastic estimations of area for land-use transition and stochastic transition algorithms. The modelling experiment presented in this paper integrates this third category, in which the transition rules are randomized, the cell transition probabilities are calculated through Bayesian probabilistic methods ('weights of evidence') and the Markov chain is, in principle, utilized for the definition of the transition rates for each possible type of land-use change. An overview of the 'weights of evidence' statistical method, as well as an explanation of how it can be applied to the modelling of urban land-use dynamics, is presented throughout the next section.

2. Methods: a Bayesian method-based cellular automaton model

2.1 *Generalization procedures applied to the land-use maps*

The city maps provided by the Bauru local authorities presented inconsistencies due to the fact that illegal settlements are not shown on official maps, and not all of the legally approved settlements drawn have, in fact, been implemented. In this way, satellite imagery arises as a feasible solution for the identification of urban settlements actually in existence, as well as for the delineation of the true urban occupation boundaries of the case study town.

In these official maps, some urban zones refer to areas which are not yet occupied and some other zones categories do not correspond to the prevailing use

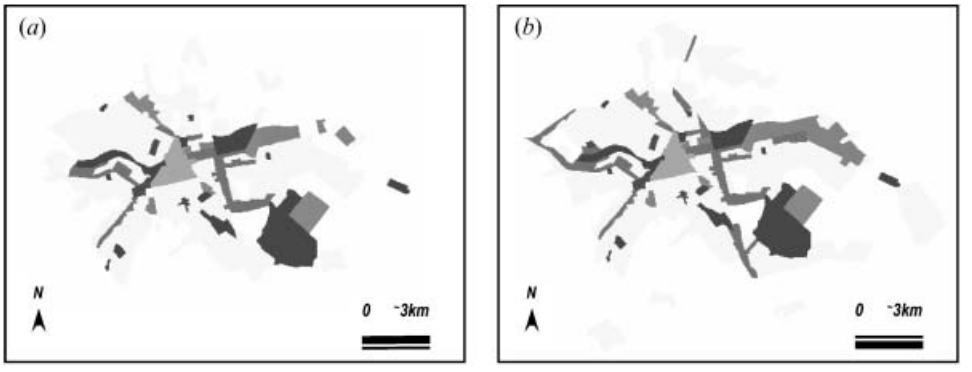

Figure 1. Land use in Bauru, São Paulo State, in (*a*) 1979 and (*b*) 1988. Residential use is light grey; institutional use is black; services use zones and corridors and industrial use are dark grey; commercial use is mid grey; and the white colour refers to non-urban use.

encountered within their limits, reflecting just the local officials intention for their future use. In this sense, the following procedures were applied to the initial (1979) and final (1988) land-use maps (figure 1) used in the simulation experiment so as to render them workable by the computational model and coherent to the reality to which they are related:

1. reclassification of the zones initially assigned by the Bauru local authorities according to their dominant and effectively existent use, based on field observations and satellite imagery;
2. reclassification of similar zones shown on official maps to only one category, e.g. residential zones of different densities were all reclassified to residential zones only; special use zones and social infrastructure equipment zones were reclassified to institutional zones only, and so on;
3. adoption of eight land-use zone categories: residential, commercial, industrial, services, institutional, mixed-use zone, leisure/recreation, and non-urban zone;
4. exclusion of districts segregated from the main urban agglomeration, i.e. those which are located above 10 km from the official urban boundary;
5. disregard of the traffic network and minor non-occupied areas in the simulations.

For updating the land-use maps used in the simulation, a Multi-Spectral Scanner (MSS) image of 22 June, 1979 (WRS 237/75) and a Thematic Mapper (TM) image of 29 November, 1988 (WRS 221/75) were employed. Official topographic charts with scale 1: 50 000 were used for the registration of the images, with a total average error of 1.308 pixel. The geographical coordinates of the control points were then used for the registration of the city maps in vector format using SPRING GIS (from the Division for Image Processing of the Brazilian National Institute for Space Research—DPI-INPE). Finally, the city maps were superimposed on linearly enhanced colour composites of the registrated images (MSS 457 and TM 147), allowing a visual cross-check of existent and non-existent settlements. It is worth remarking that if modelling is concerned only with urban

expansion, i.e. only the classes of urban and non-urban use are under consideration, detailed land-use maps derived from aerial photos and field checking are no longer necessary and a simple binary classification of remotely sensed data suffices.

2.2 *Exploratory analysis and selection of variables*

As stated formerly, the maps of explaining variables relate to the technical infrastructure and socio-economic aspects of Bauru. Initially, these maps were scanned in the OCE scanner (model G6035S) and digitized in AutoCad release 2000. They were then exported as files with extension DXF to SPRING. These same procedures were also adopted for the production of the Bauru land-use maps presented in figure 1. In SPRING, the maps of variables were then subjected to a preliminary processing, including vector edition, polygons identification, elaboration of distance maps and spatial statistical analysis maps, such as the Kernel points density estimator, etc.

Since the 'weights of evidence' statistical method (to be employed in the calculation of the cells' transition probabilities) is based on the 'Bayes theorem of conditional probability', the selection of variables for the modelling analysis ought to take into account the checking of independence amongst pairs of variables chosen to explain the same category of land-use change.

For this end, two methods were used: the Cramer's Coefficient (V) and the Joint Information Uncertainty (U). In both cases, it is necessary to obtain values from an area cross-tabulation between pairs of maps of variables under analysis. Let the area table between map A and map B be called matrix $\mathbf{T}$, with elements $\mathbf{T}_{ij}$, where there are $i=1, 2, \ldots, n$ classes of map B (rows of the table) and $j=1, 2, \ldots, m$ classes of map A (columns of the table). The marginal totals of $\mathbf{T}$ are defined as $\mathbf{T}_{i.}$ for the sum of the i-th row, $\mathbf{T}_{.j}$ for the sum of the j-th column, and $\mathbf{T}_{..}$ for the grand total summed over rows and columns. If the two maps are independent of one another, with no correlation between them, then the expected area in each overlap category is given by the product of the marginal totals, divided by the grand total. Thus, the expected area $\mathbf{T}^*_{ij}$ for the i-th row and j-th column is:

$$\mathbf{T}^*_{ij} = \frac{\mathbf{T}_{i.}\mathbf{T}_{j}}{\mathbf{T}_{..}}. \tag{1}$$

Then, the chi-square statistic is defined as:

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\left(\mathbf{T}_{ij} - \mathbf{T}^*_{ij}\right)^2}{\mathbf{T}^*_{ij}}, \tag{2}$$

the familiar (observed$-$expected)^2/expected expression, which has a lower limit of 0 when the observed areas exactly equal the expected areas, and the two maps are completely independent. The Cramer's Coefficient (V) is then defined as:

$$V = \sqrt{\frac{\chi^2}{\mathbf{T}.M}}, \tag{3}$$

where M is the minimum of ($n-1$, $m-1$).

The Joint Information Uncertainty (U) belongs to the class of entropy measures, which are also based on the area cross-tabulation matrix **T**, but can also be used for measuring associations. Suppose that the $\mathbf{T}_{ij}$ values are transformed to area proportions, p, by dividing each area element by the grand total $\mathbf{T}_{..}$. Thus, $p_{ij}=\mathbf{T}_{ij}/\mathbf{T}_{..}$, and the marginal proportions are defined as $p_{i.}=\mathbf{T}_{i.}/\mathbf{T}_{..}$ and as $p_{.j}=\mathbf{T}_{.j}/\mathbf{T}_{..}$. Therefore entropy measures, also known as information statistics, can be defined using the area proportions as estimates of probabilities. Proportions are dimensionless, so entropy measures have the advantage over chi-squared measures of being unaffected by measurement units (Bonham-Carter 1994).

Assuming that an area proportions matrix for map A and map B has been determined from **T**, then the entropy of A and B is defined as:

$$H(A)=-\sum_{j=1}^{m} p_j-\ln p_j \tag{4}$$

$$H(B)=-\sum_{i=1}^{n} p_i-\ln p_i\ , \tag{5}$$

where ln is the natural logarithm. The joint entropy of the combination, $H(A,B)$, is simply

$$H(A,B)=-\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\times \ln p_{ij}\ . \tag{6}$$

Then the 'Joint Information Uncertainty' of A and B, $U(A,B)$, can be used as a measure of association and is defined as

$$U(A,B)=2\left[\frac{H(A)+H(B)-H(A,B)}{H(A)+H(B)}\right] \tag{7}$$

which varies between 0 and 1. When the two maps are completely independent, then $H(A,B)=H(A)+H(B)$ and $U(A,B)$ is 0, and when the two maps are completely dependent, $H(A)=H(B)=H(A,B)=1$, and $U(A,B)$ is 1.

The criterion which is used to determine whether one variable is independent of another is, to a large extent, arbitrary as there is no large body of case results associated with the application of these methods. Where this particular variant of logit modelling has been used in the geosciences, Bonham-Carter (1994) reports that values less than 0.5 for Cramer's Coefficient and the Joint Information Uncertainty suggest less association rather than more. In all comparisons made here, these associations are less than this threshold. Indeed, all values are less than 0.45 for V and less than 0.35 for U.

The variables employed in the simulations and their notations are found in table 1 and the respective values for the measures of association (V and U) between pairs of variables chosen to explain the same category of land-use change are found in table 2. As none of the association measure values surpassed the thresholds, no variables that were preliminarily selected for modelling have been discarded from the analysis. In practice, the variables selection routine also includes empirical procedures, based on the visualization of distinct variables superposed on the final land-use map, so as to identify those more meaningful to explain the different types of land-use change (figure 2).

Table 1. Definition of the 12 independent land-use change evidences or variables.

Notation	Physical or socio-economic land-use change evidence
water	Area served by water supply.
mh_dens	Medium-high density of occupation (25–40%).
soc_hous	Existence of social housing.
com_kern	Distances to different ranges of commercial activities concentration, defined by the Kernel estimator.
dist_ind	Distances to industrial zones.
dist_res	Distances to residential zones.
per_res	Distances to peripheral residential settlements, isolated from the urban concentration.
dist_inst	Distances to social infrastructure (institutional use), isolated from the urban concentration.
exist_rds	Distances to main existing roads.
serv_axes	Distances to the services and industrial axes.
plan_rds	Distances to planned roads.
per_rds	Distances to peripheral roads, which pass through non-occupied areas.

2.3 *Estimation of global transition rates*

For the specific case study town in question—Bauru—in the period 1979–1988, five types of land-use change were detected (table 3). In order to calculate global land-use transition rates for the period 1979–1988, the initial and final land-use maps were converted to raster files with extension TIFF and resolution 100×100 (m), and then exported to the IDRISI Geographical Information System. The adopted resolution is about a city block size, which was deemed convenient for the purpose of urban land-use analysis, for intra-city block variations in land use are disregarded. In IDRISI, a cross-tabulation operation was made between both land-use maps (see

Table 2. Associations between the independent evidences (variables).

Evidence A	Evidence B	Cramer's Statistic $V_{A,B}$	Uncertainty $U_{A,B}$
water	*serv_axes*	0.3257	0.0767
mh_dens	*soc_hous*	0.0460	0.0017
	plan_rds	0.2617	0.0701
	per_rds	0.0201	0.0003
soc_hous	*plan_rds*	0.1174	0.0188
	per_rds	0.0480	0.0047
com_kern	*dist_res*	0.4129	0.3447
	per_res	0.1142	0.0310
	dist_inst	0.1218	0.0520
	exist_rds	0.2685	0.1499
	serv_axes	0.2029	0.1099
	per_rds	0.0434	0.0064
dist_ind	*serv_axes*	0.1466	0.0477
dist_res	*serv_axes*	0.2142	0.1002
per_res	*dist_inst*	0.1487	0.0559
	exist_rds	0.0592	0.0078
	per_rds	0.1733	0.0553
dist_inst	*exist_rds*	0.0601	0.0108
	per_rds	0.0765	0.0238
exist_rds	*per_rds*	0.0239	0.0019
plan_rds	*per_rds*	0.0247	0.0029

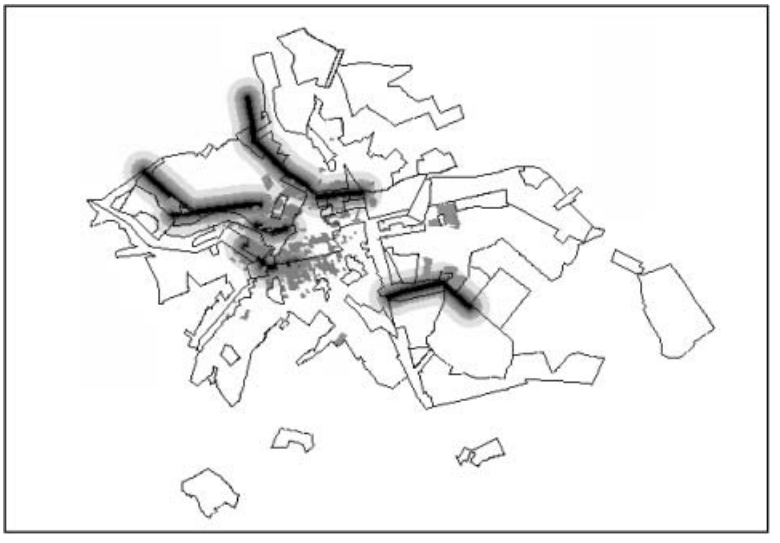

Figure 2. Spatial independence of factors determining the transition from residential to mixed use (RES_MIX). The buffer bands are distance to planned roads (plan_rds), the darker diffused spots are areas of medium–high density of occupation (mh_dens) and the greater dark polygons correspond to social housing (soc_hous).

figure 1) so as to generate global transition percentages (table 4) for the five existing types of land-use change.

For the estimation of global transition percentages in the case of simulations where the final land-use map was unavailable or in the case of land-use change forecasts through DINAMICA, the Markov chain ought to be employed. This

Table 3. Existing 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

Table 4. The matrix of transition rates for Bauru, 1979–1988.

Land use	*NonU*	*Res*	*Comm*	*Indust*	*Inst*	*Serv*	*Mixed*	*Leis/Rec*
NonU	0.9171	0.0698	0	0.0095	0	0.0036	0	0
Res	0	0.9380	0	0	0	0.0597	0.0023	0
Comm	0	0	1	0	0	0	0	0
Indust	0	0	0	1	0	0	0	0
Inst	0	0	0	0	1	0	0	0
Serv	0	0	0	0	0	1	0	0
Mixed	0	0	0	0	0	0	1	0
Leis/Rec	0	0	0	0	0	0	0	1

chain is a mathematical model designed to describe a certain type of process that moves in a sequence of steps through a set of states, whose formula is defined as:

$$\pi(t+1) = P\pi(t) \,, \tag{8}$$

where $\pi(t)$ is a column vector, with n elements, that represents the system condition in a certain time t (e.g. area percentages for each n_i land-use category or state); $\pi(t+1)$ is the vector representing the occupation of n states in a given future time $t+1$; and P is the transition probability matrix or the table for land-use transition rates.

An important constraint of the Markov model lies in the fact that, in principle, it supposes that transition probabilities do not change over time (stationary process). Moreover, given its stochastic nature, the Markov chain masks the causative variables. It is not an explanatory model and is, thus, of no use in understanding the causes and driving factors of land-use transition processes. On the other hand, the Markov chain analysis has the great advantage of presenting a mathematical and operational simplicity. Simple trend projection involves no more than matrix multiplication and the only data requirement is for current land-use information (JRC and ESA 1994).

2.4 *Reckoning of the cells' land-use transition probabilities*

As previously said, the 'weights of evidence' statistical method, employed in the calculation of the cells' transition probabilities, is based on the 'Bayes theorem of conditional probability'. Basically, this theorem concerns the favourability of detecting a certain event, which, in the current case, can be a given category of land-use change (e.g. non-urban use to residential use), provided that an evidence (e.g. water supply area), also called explaining variable, has already happened. The evidences or explaining variables of the experiment presented in this paper, which are summarized in table 1, mainly refer to the technical infrastructure and socio-economic aspects of the case study town, Bauru.

The favourability to find the event (change from non-urban to residential use) R, given the presence of the evidence (water supply) S, can be expressed by:

$$P\{R/S\} = \frac{P\{R \cap S\}}{P\{S\}} \,, \tag{9}$$

where $P\{R/S\}$ is the conditional probability of the event R occurring, given the presence of the explaining variable S. The equations of the Bayes theorem can be expressed in an odds form. Odds are defined as a ratio of the probability that an event will occur to the probability that it will not occur. The weights of evidence method uses the natural logarithm of odds, known as log odds or logits. In this way, through some algebraic manipulations, the following expression is obtained:

$$O\{R/S\} = O\{R\} \frac{P\{S/R\}}{P\{S/\overline{R}\}} \,, \tag{10}$$

where $O\{R/S\}$ is the conditional (posterior) odds of R given S, $O\{R\}$ is the prior odds of R and $P\{S/R\}/P\{S/\overline{R}\}$ is known as the sufficiency ratio (LS). In weights of evidence, the natural logarithm of both sides of equation (10) is taken, and $\log_e LS$ is the positive weight of evidence W^+, which is calculated from the data. Then:

$$\log\text{it}\{R/S\} = \log\text{it}\{R\} + W^+ \,. \tag{11}$$

Similarly, the logits expression for the conditional probability of R, given the absence of the evidence $\overline{S}$, will provide the negative weight of evidence W^-:

$$\log it\{R/S\} = \log it\{R\} + W^- . \tag{12}$$

If the evidence is uncorrelated with the events, then $LS=1$, and the posterior probability equals the prior probability, and the probability of an event would be unaffected by the presence or absence of a certain evidence. On the other hand, W^+ is positive, and W^- is negative, due to the positive correlation between the evidences and the events. Conversely W^+ would be negative and W^- positive for the case where a very limited part of the event occurs on the evidence area than would be expected due to chance. Thus, if the events are independent of whether the evidence is present or not, then $W^+ = W^- = 0$. In the weights of evidence method, there is a specific way for calculating probability ratios (odds) in the case of n maps of evidence or variables (V_i). The general expression for combining $i=1, 2, \ldots, n$ maps is either:

$$O\{R/V_1, V_2, V_3, ..., V_n\} = O\{R\} \prod_{i=1}^{n} LS_i \, , \tag{13}$$

for the likelihood ratios or:

$$\log it\{R/V_1, V_2, V_3, ..., V_n\} = \log it\{R\} + \sum_{i=1}^{n} W^+ \, , \tag{14}$$

for the weights. In the particular case of the DINAMICA simulation model, adopted for the modelling experiment being considered, the cells' transition probabilities are calculated through a formula which converts logit into conditional probability, as follows:

$$P_{x,y}\{R/V_1, ..., V_n\} = \psi \frac{O\{R\} e^{\sum_{i=1}^{n} W^+_{x,y}}}{1 + O\{R\} e^{\sum_{i=1}^{n} W^+_{x,y}}} \tag{15}$$

where V refers to all possible variables (evidences) selected to explain the transition R and ψ corresponds to a normalizing constant, required to ensure that the conditional probability of all cells with coordinates x,y lies between 0 and 1.

The first step in the process of calculating the cells' transition probabilities using DINAMICA is to obtain a cross-tabulation map (figure 3) between the initial and final land-use maps elaborated for the city of Bauru, respectively, for the years 1979 and 1988.

In IDRISI, the land-use cross-tabulation map of Bauru (1979–1988) was used to generate land-use transition maps (seen on the right column of figure 4) for each of the five possible types of land-use change presented in table 3. This was done through reclassification tables ('edit' command), on which three basic rules were observed. First, all raster values corresponding to classes of land-use permanence or transition whose initial land use was different from the initial land-use category in the considered type of land-use change were assigned value 0 (black colour). This reclassification to value 0 is automatic for raster values not included in the 'edit' table. Secondly, all raster values corresponding to classes of land-use transition whose initial and final land-use categories were equal to the initial and final categories of the land-use change at issue were assigned value 2 (dark grey). Thirdly, all other remaining classes of land-use permanence or transition were assigned value 1 (light grey).

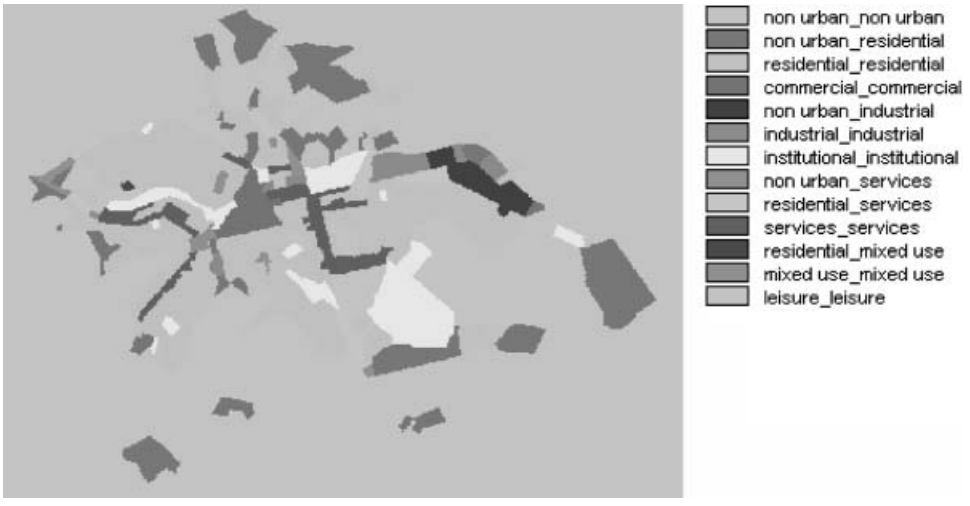

Figure 3. Land-use change 1979 to 1988.

Once all possible types of land-use transition maps were elaborated (nu_res; nu_ind; nu_serv; res_serv; res_mix), they were then subjected to partial cross-tabulations with the respectively selected evidences maps. The sets of evidences maps selected to explain each of the five types of land-use transitions are found in table 5. The evidences maps, pre-processed in the SPRING GIS, were converted to raster files—in the same manner as the initial and final land-use maps—with extension TIFF and resolution 100×100 (m) and then exported to IDRISI. The partial cross-tabulations disregard the raster values 0 (black colour) in the land-use transition maps and are accomplished through the 'ermatt' command of IDRISI.

The numerical values of cell proportions existing in the absence/presence of a binary evidence (e.g. water supply) or in the different ranges of distances maps and found to be overlying on either class 1 or 2 of the land-use transition maps are (for each cross-tabulation operation) selectively transferred to EXCEL files specially created for the calculation of the weights of evidence, as shown in equations (10) and (11).

Using the values for the positive weights of evidence W^+ concerning the several evidences maps employed in the analysis of each category of land-use change (table 6), the DINAMICA simulation model will then calculate the cells' transition probabilities according to equation (15) for the five types of land-use transition. By means of the cells' transition probabilities, DINAMICA will generate the respective transition probability maps (seen on the left column of figure 4) for each of the five types of land-use change existing in Bauru from 1979 to 1988. These maps are seen in ERMAPPER, a GIS employed by DINAMICA for visualization purposes.

It is worth noticing how well these probability maps detect the transition areas (dark grey) on the corresponding land-use transition maps, for all the light and

Figure 4. Estimated transition probability surfaces and land use change, 1979–1988. In the left column, the range of probabilities runs from high (mid grey) and medium (light grey) to low (dark grey) and null (black). In the right column, changes in land use are in dark grey, land-use permanences are in light grey, whereas areas disregarded for the considered type of land-use transition are in black.

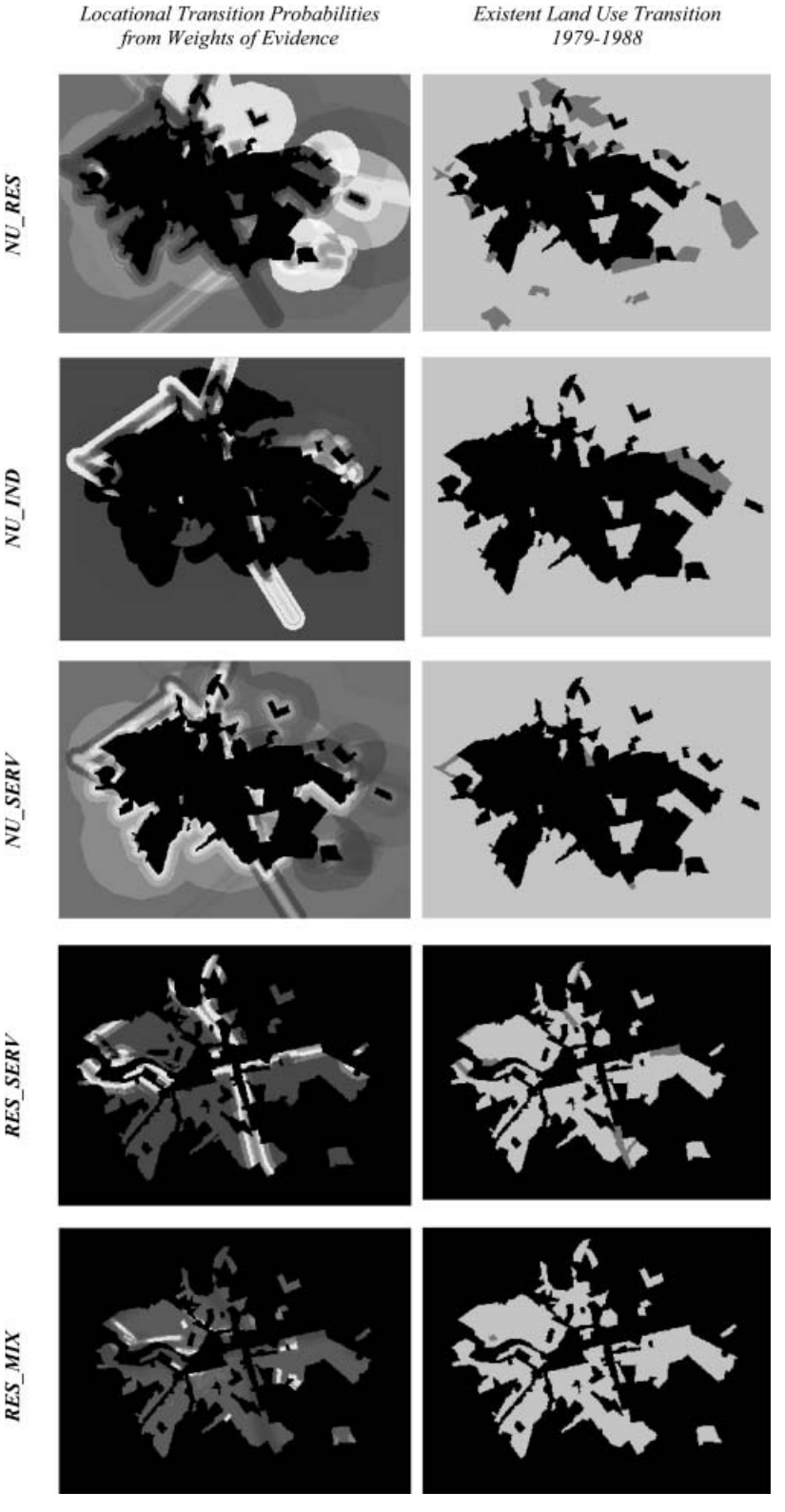
Locational Transition Probabilities from Weights of Evidence
Existent Land Use Transition 1979-1988
NU_RES
NU_IND
NU_SERV
RES_SERV
RES_MIX

Table 5. Selection of evidences (variables) determining land-use change.

Evidences	*NU_RES*	*NU_IND*	*NU_SERV*	*RES_SERV*	*RES_MIX*
water				•	
mh_dens					•
soc_hous					•
com_kern	•		•		
dist_ind		•			
dist_res			•		
per_res	•				
dist_inst	•				
exist_rds	•				
serv_axes		•	•	•	
plan_rds					•
per_rds	•				•

Table 6. The weights of evidence.

Land-use transitions	Evidences	Positive weights of evidence $W^{+}_{x,y}$						
		1	*2*	*3*	*4*	*5*	*6*	*7*
NU_RES	*com_kern**	3.749	2.106	1.864	0.491	−0.323	0	na
	per_res‡	1.968	1.615	1.392	0.892	−0.626	−0.469	na
	dist_inst§	0.003	0.600	1.254	0.727	−0.359	−0.089	na
	exist_rds¶	0.231	0.320	0.353	0.510	0.443	0.196	−0.085
	per_rds‖	2.377	2.269	2.068	1.984	1.444	0.857	−0.127
NU_IND	*dist_ind*†	3.862	4.016	3.792	3.452	1.763	0	0
	serv_axes¶	2.722	2.799	2.676	2.625	2.525	1.727	−3.832
NU_SERV	*com_kern**	3.412	4.469	2.912	0.878	0	0	na
	dist_res‡	2.144	1.523	0.621	−0.065	0	0	na
	serv_axes¶	3.508	3.321	2.917	1.869	0.450	0	0
RES_SERV	*water*	*Presence* −0.6611				*Absence* 0.2883		
	serv_axes¶	2.780	1.948	1.461	0.888	−0.297	−1.412	−3.284
RES_MIX	*mh_dens*	*Presence* 0.6452				*Absence* −0.0635		
	soc_hous	*Presence* 2.4678				*Absence* −0.3214		
	plan_rds¶	3.506	1.863	0	0	0	0	0
	per_rds‖	1.775	1.652	1.848	0.903	0	0	0

Note: Distance bands in metres:
*1: 0–500; 2: 500–1000; 3: 1000–1500; 4: 1500–10 000; 5: 10 000–30 000; 6: >30 000
†1: 0–500; 2: 500–1000; 3: 1000–1500; 4: 1500–2000; 5: 2000–5000; 6: 5000–10 000; 7: >10 000
‡1: 0–500; 2: 500–1000; 3: 1000–2000; 4: 2000–5000; 5: 5000–10 000; 6: >10 000
§1: 0–500; 2: 500–1000; 3: 1000–3000; 4: 3000–8000; 5: 8000–15 000; 6: >15 000
¶1: 0–250; 2: 250–500; 3: 500–750; 4: 750–1000; 5: 1000–1250; 6: 1250–2000; 7: >2000
‖1: 0–250; 2: 250–500; 3: 500–750; 4: 750–1000; 5: 1000–1500; 6: 1500–2500; 7: >2500

mid-grey regions on the probability maps relate to the very areas owning the highest transition probability rates.

2.5 *Model calibration*

For the calibration of the model, empirical procedures were adopted. They basically concern the visual comparative analysis for each type of land-use change amongst the general trends of preliminary simulation results, the hints provided by both the

transition probability map and the land-use transition map and the guideline information contained in the simultaneous overlay of different explaining variables maps upon the final land-use map in vector format.

The model calibration, on the other hand, is as well accomplished by the analysis of scatter plots relating subcategories of evidences (distances ranges)—whenever they are available—with their respective positive weights of evidence. In a general manner, when the plots present a good fit of trend lines (which can assume different orders and types), the evidences to which they are associated are highly likely to be included in the model.

The final decision towards the inclusion or exclusion of a given evidence will always rely upon a broad judgement, in which the environmental importance of the evidence and its coherence concerning the phenomenon (land-use transition) being modelled are analysed. As stated by Couclelis (1997), to take full advantage of CA models as simulating (and forecasting) tools, planners and others need to rely as much on their right-brain powers of pattern recognition and relationship perception as on left-brain analyses of the inevitably inaccurate quantitative outputs.

3. Results and discussion

Upon the basis of the calibration process carried out, it becomes evident that the probability of certain non-urban areas in the city of Bauru to shelter residential settlements ('nu_res' land-use transition) depends largely on the previous existence of this type of settlement in the surroundings, on the greater proximity of these areas to commercial activities clusters as well as on the available accessibility to such areas.

As to the transition of non-urban areas to industrial use (nu_ind), there are two great driving forces: the nearness of such areas to the previously existing industrial use and the availability of road access. This can be explained by the fact that in the industrial production process, the output of certain industries represents the input of other ones, which raises the need of rationalization and optimization of costs by the clustering of plants interrelated in the same production chain. Furthermore, plots in the vicinities of industrial areas are often prone to be devalued for other uses, which makes them rather competitive for the industrial use.

Regarding the transition of non-urban areas to services use (nu_serv), three major factors are crucial: the proximity of these areas to clusters of commercial activities; their closeness to areas of residential use; and, last but not least, their strategic location in relation to the N–S/E–W services axes of Bauru. The first factor accounts for the suppliers' market (and, in some cases, also consumers' market) of services; the second factor represents the consumers' market proper; and the third factor corresponds to the accessibility for both markets related to the services use.

Since the transition 'residential to services use' (res_serv) already takes place relatively close to the suppliers and consumers markets, it will solely consider a strategic location in relation to the N–S/E–W services axes of Bauru, as well as the absence of water supply at the initial time of simulation (note in table 6 that W^+ in this particular case is negative). Localities close to the services axes but deprived of water supply refer to the immediate fringe of consolidated urban areas, i.e. newly developed areas, will be integrated fully into the urban network by the final time of simulation.

Finally, the last type of land-use transition concerns the shift from residential use to mixed use (res_mix). The mixed-use zones, which actually play the role of urban

sub-centres, constitute a strengthening of minor commercial centres, which at a later stage also start to attract services and social infrastructure equipment besides more diversified commercial activities. Therefore, new mixed-use zones arise in more peripheral areas, where a greater occupational gathering is ensured. Thus, the decisive factors for this last type of land-use change are:

- the existence of medium–high density of occupation (higher density values only occur in the central commercial zone of the town or in the immediacies of already existing mixed-use zones);
- the presence or proximity of social housing settlements (for they shelter the greatest occupational densities in more peripheral areas and, hence, greater consumers' markets);
- the nearness to planned or peripheral roads, since new mixed-use zones arise in further areas of the town.

It is implied by the above analysis that the land-use transitions show compliance with economic theories of urban growth and change, where there is a continuous search for the optimal location, able to ensure competitive real estate prices, good accessibility, rationalization of transportation costs, and a strategic location in relation to markets.

After the calibration for the selection of evidences maps sets is accomplished, a new calibration process concerning the script parameters of the DINAMICA simulation model takes place. Such parameters refer to the number of iterations (runs), proportion of cell transitions by contiguity ('expander' operator) and by nucleation ('patcher' operator), average size and variance of patches to be generated either by the expander or patcher operators, etc.

The expander is an algorithm of the DINAMICA model which realizes transitions from a state i to a state j only in the adjacent vicinities of cells with state j. The patcher operator, in turn, accomplishes transitions from a state i to a state j only in the adjacent vicinities of cells with state other than j.

Due to the randomness of the DINAMICA transition algorithms, even though the same sets of evidences maps for each type of land-use transition and the same script parameters are kept in different runs, distinct simulation results will be produced after each run of the model. In this way, the three best urban land-use simulation results for the city of Bauru in the period 1979–1988 can be presented (figure 5).

The transition 'non-urban–residential use' proved to be the most challenging for simulation. Its boundaries are defined by highly unstable factors, such as the real estate entrepreneurs' decisions to develop certain areas rather than others, and the plot boundaries themselves, which can have their forms drastically altered by merging or splitting operations.

The services corridors (light mid-grey) were well modelled in all simulations. The industrial use zone (mid-grey in the north-eastern part of town) was considerably well detected in all of the three simulations results, especially in S2 and S3. The leisure and recreation zones (very dark grey), the institutional zones (very light grey) and the central commercial zone (dark mid-grey triangle in the town centre) did not suffer any transitions. The new mixed zone that arose in the north-western part of the town during the simulation period was rather well modelled, particularly in S1 and S3.

To conclude, it is worth stressing here the wide feasibility (and the cells' transition probability maps are a concrete proof for this) to optimize the simulation results by means of a model which embraces more refined algorithmic logics, such as the incorporation of fractal parameters in the transition algorithms as well as the possibility to define patch average sizes and variances for the expander and patcher algorithms separately.

4. Statistical validation of the model

With the purpose of conducting statistical tests for the spatial validation of models of land-use dynamics, Constanza (1989) presented a procedure entitled 'Multiple Resolution Method', in which a sampling window, which can assume different sizes, moves over the entire image considered, and the average fit between two given scenes (the real and the simulated one) for a particular window size is calculated. In this estimation, a comparative analysis is conducted between the absolute number of pixels belonging to the same classes existing on both scenes and found within a given window. This multiple resolution method was implemented in a UNIX environment program named FIT, developed by CSR-UFMG. FIT was applied for the best simulation results presented in figure 5, with sampling window sizes of 3×3, 5×5 and 10×10, and the values for goodness-of-fit obtained were 0.902937, 0.896092 and 0.901134, respectively, for S1, S2 and S3.

5. Conclusions

The urban land-use dynamics models, driven by GIS and remote sensing data, have proved to be useful for the identification of main urban growth vectors and their general land-use tendencies, which enables local planning authorities to manage and reorganize (if it comes into question) city growth according to the environmental carrying capacity of concerned sites and to their present and envisaged infrastructure availability.

The urban expansion forecasts provided by such models also help local authorities in general to establish investment goals in terms of technical and social infrastructure equipment. Decision-makers from the private sphere can also benefit from the modelling output data, since transportation, conventional and mobile phones, cable TV and internet companies, among others, will have subsidies for defining priorities as to where and how intensely to invest. In addition, organized civil society, either through NGOs or local associations, can profit from the modelling forecasts in order to enhance, by legal means, demanding social movements for the implementation of social and technical infrastructure. Their requests and respective arguments will be based on realistic short- and medium-term urban growth trends.

Finally, it is worth mentioning that the 'weights of evidence' statistical method is not constrained by the straitjacket of rigid theory devices; neither does it impose theoretical restraints on the modelling objects. Since this a wholly empirical approach, its applicability can be extended to further cities—Brazilian and world-wide—provided that the minimum necessary sets of evidences maps are available.

Acknowledgements

The authors wish to thank the Water Supply and Waste Water Disposal Department (DAE), and the Bauru Planning Secretariat (SEPLAN) for providing

the city maps. They are also grateful for the help and co-operation of the technical and administrative staff of the Centre for Remote Sensing of the Federal University of Minas Gerais (CSR-UFMG). Crucial financial support was granted to this research by the Academic Coordination for Remote Sensing of the Brazilian National Institute for Space Research (PG-SER/INPE), the São Paulo State Foundation for Research Support (FAPESP) and the Brazilian National Foundation for the Undergraduates and Graduates Upgrade (CAPES).

References

BATTY, M., 2000, GeoComputation using cellular automata. In *Geocomputation*, S. Openshaw and R.J. Abrahart (Eds) (New York: Taylor & Francis), pp. 95–126.

BATTY, M. and XIE, Y., 1994, From cells to cities. *Environment and Planning B*, **21**, pp. 31–48.

BATTY, M. and XIE, Y., 1997, Possible urban automata. *Environment and Planning B*, **24**, pp. 175–192.

BENATI, S., 1997, A cellular automaton for the simulation of competitive location. *Environment and Planning B*, **24**, pp. 205–218.

BONHAM-CARTER G.F., 1994 *Geographic Information Systems for Geoscientists: Modelling with GIS* (Ontario: Pergamon).

CLARKE, K.C., HOPPEN, S. and GAYDOS, L., 1997, A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, **24**, pp. 247–261.

CONSTANZA, R., 1989, Model goodness of fit: a multiple resolution procedure. *Ecological Modelling*, **47**, pp. 199–215.

COUCLELIS, H., 1997, From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B*, **24**, pp. 165–174.

DEADMAN, P., BROWN, R.D. and GIMBLETT, P., 1993, Modelling rural residential settlement patterns with cellular automata. *Journal of Environment Management*, **37**, pp. 147–160.

JRC (Joint Research Centre—European Commission/Institute for Remote Sensing Applications) and ESA (European Space Agency/ESRIN—Earthnet Programme Office), 1994, *Modelling Deforestation Processes—A Review*, Trees Series B (Luxembourg: ECSC-EC-EAEC, Research Report no.1).

PHIPPS, M. and LANGLOIS, A., 1997, Spatial dynamics, cellular automata, and parallel processing computers. *Environment and Planning B*, **24**, pp. 193–204.

PORTUGALI, J., BENENSON, I. and OMER, I., 1997, Spatial cognitive dissonance and sociospatial emergence in a self-organizing city. *Environment and Planning B*, **24**, pp. 263–285.

WHITE, R.W. and ENGELEN, G., 1997, Cellular automaton as the basis of integrated dynamic regional modelling. *Environment and Planning B*, **24**, pp. 235–246.

WHITE, R.W., ENGELEN, G. and ULJEE, I., 1998, *Vulnerability Assessment of Low-Lying Coastal Areas and Small Islands to Climate Change and Sea Level Rise – Phase 2: Case Study St. Lucia* (Kingston, Jamaica: RIKS Publication, Report to the United Nations Environment Programme, Caribbean Regional Co-ordinating Unit).