1

An Interactive Method to Dynamically Create Transition Rules in a Land-use Cellular Automata Model

Hasbani, J.-G., N. Wijesekara and D.J. Marceau
Department of Geomatics Engineering,
University of Calgary
Canada

1. Introduction

Cellular automata (CA) models are increasingly applied to simulate a wide range of spatio-temporal phenomena, including urban traffic (Sun and Wang, 2007), fire propagation (Ohgai *et al.*, 2007), and insect infestation (Bone *et al.* 2006), but most importantly urban development (Almeida *et al.*, 2008; Benenson and Torrens, 2004; Clarke *et al.*, 1997; Santé *et al.*, 2010; Shen *et al.*, 2009; Van Vliet *et al.*, 2009), and land-use changes (Ménard and Marceau, 2007; Moreno *et al.*, 2010; Soares-Filho *et al.*, 2002; Sui and Zeng, 2001). CA models are particularly suitable for land-use change modeling for several reasons. They are explicitly spatial and can be constrained in various ways to reflect local tendencies (Jenerette and Wu, 2001; Li and Yeh, 2000). It is also possible to specify for each simulated time step the quantity of land that should change from one land use to another (Jantz and Goetz, 2005). Information from a-spatial models, like a population growth model, can be integrated into the CA model to spatially allocate the land-use changes (White *et al.*, 1997). A stochastic factor can also be included in the model to take into account some degree of unpredictability in the system (Moreno *et al.*, 2009). As a consequence, CA models are often designed to test what-if scenarios and policies in urban and regional planning (Erlien *et al.*, 2006; Jantz *et al.*, 2003; Li and Yeh, 2004).

However, a challenge when implementing a CA model is its calibration. Calibration involves finding the parameters of the transition rules and the numerical values of these parameters so that the rules closely correspond to the dynamics of the system under investigation. This process is complicated due to the large number of combinations involved when several cell states, state transitions, parameters, and parameter values are being considered (Li and Yeh, 2002a; Shan *et al.*, 2008). In addition, such combinations do not necessarily yield unique solutions (Verburg *et al.*, 2004). Since there is no obvious way of finding which parameter should or should not be included in the model, the transition rules are often based on the modeler's intuitive understanding of the driving factors affecting the system (Wu, 2002).

Statistical techniques, such as logistic and multiple regressions (Fang *et al.*, 2005; Sui and Zeng, 2001; Wu, 2002), principal component analysis (Li and Yeh, 2002a), and multivariate analysis of variance (Lau and Kam, 2005) have been proposed for CA calibration. Computational intelligence techniques have also been tested, including artificial neural network (Li and Yeh, 2002b; Pijanowski *et al*., 2002), genetic algorithm (Shan *et al*., 2008), and data mining (Wang *et al*., 2010). Other methods involve the systematic testing of parameters (Jantz and Goetz, 2005; Jantz *et al.*, 2003) and iterative calibration to achieve reasonable goodness-of-fit (Straatman *et al*., 2004). While these approaches might provide satisfactory simulation results, they often leave the modeler with little control on the mathematical equations used to determine the transition rules and the difficulty of understanding the geographical meaning of these rules (Verburg *et al*., 2004).

This paper describes a semi-automated, interactive method that was designed and implemented to dynamically create transition rules and calibrate a land-use CA model. The proposed method combines the benefits of conditional and mathematical rules and is adaptable in terms of number of land-use classes, and spatial and temporal scale of the input data. It allows the modeler to acquire information about the importance of the factors associated to historical land-use changes within the study area and to interactively select the parameter values required for the model calibration. A detailed description of the steps involved in the CA calibration is provided. The CA model is then used to answer the following questions: a) how sensitive is the model to the conditions involved in the calibration, including the cell size, neighborhood configuration, parameter values and external driving factors? b) what is the performance of the model, in terms of presence and location, in simulating land-use changes using the transition rules identified by the proposed calibration method?

2. Methodology

The study area is the dynamic eastern portion of the Elbow River watershed, located in southern Alberta, Canada, that covers an area of about 600 km^2 (Figure 1). The area is experiencing considerable pressure for land-use development due to the booming of the Alberta economy and its proximity to the City of Calgary, a fast growing city of one million inhabitants. About 5% of the watershed lies within the City of Calgary; 10% lies within the Tsuu T'ina nation, 20% within the municipal district of Rocky View, and the remaining 65% within the Kananaskis country. The study area is covered by about 48% of forest, 40% of agriculture and grassland, and 10% of built-up areas.

The historical land-use maps required for the CA calibration and validation were generated from Landsat Thematic Mapper imagery acquired during the summers of 1985, 1992, 1996, 2001, 2006 and 2010 at the spatial resolution of 30 m. Seven dominant classes were identified, namely evergreen, deciduous, agriculture, rangeland and parkland, built-up areas, water and clear-cut. Field verification was conducted for the years 2006 and 2010 and ancillary data along with expert knowledge were used to verify the classification results. A computer program was developed and applied to identify and correct minor spatial-temporal inconsistencies due to classification and georeference errors in the historical land-use maps.

A graph of the historical land-use trends reveals a decrease in the forested areas, a slight increase in parkland/rangeland, a sharper increase of built-up areas while agriculture slightly fluctuates, mostly from 2002 (Figure 2).

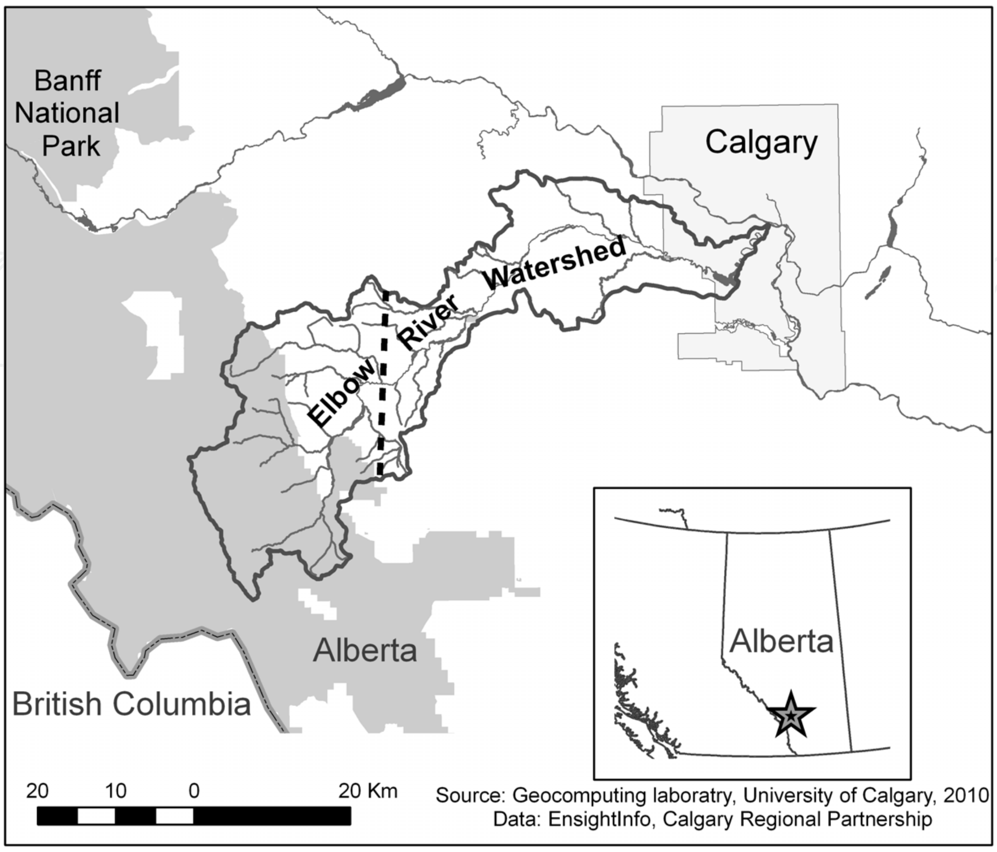


Fig. 1. Location of the study area; the dashed line represents the western limit of the study area

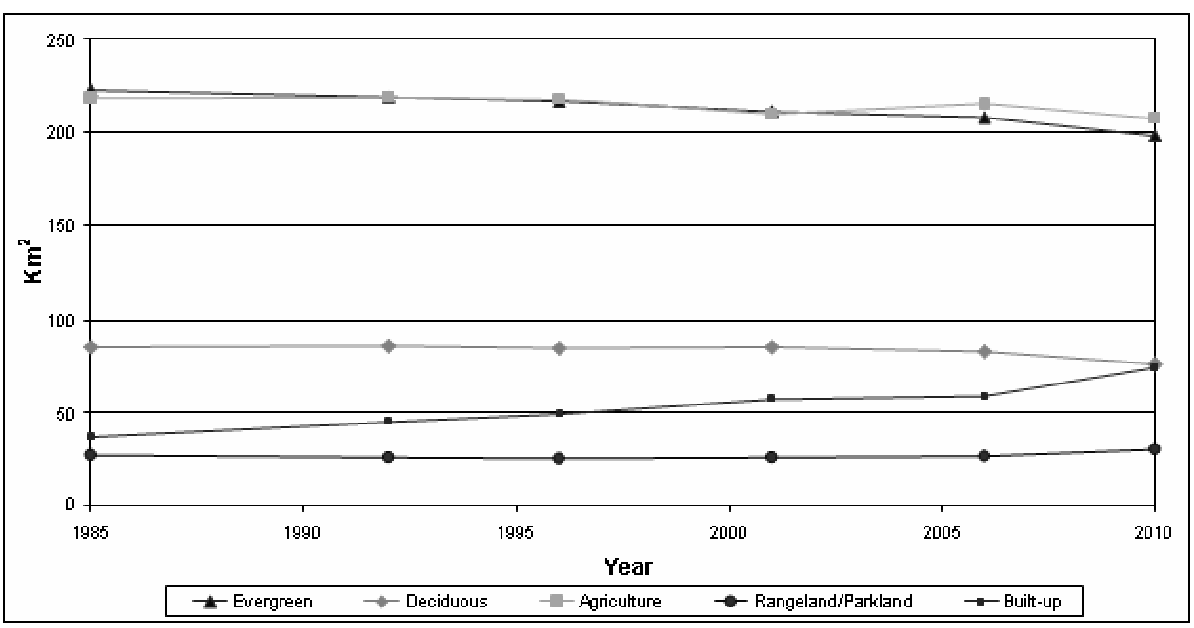


Fig. 2. Historical land-use trends in the study area

The historical land-use maps also indicate that a considerable amount of land-use transition occurred in the study area during the period considered (Table 1).

From	To	Land-use transition (%)	Total (%)
Evergreen	Agriculture	6.23	14.47
Deciduous		6.41	
Rangeland/Parkland		1.83	
Evergreen	Built-up	11.40	98.16
Deciduous		17.88	
Agriculture		65.97	
Rangeland/Parkland		2.91	
Agriculture	Rangeland /Parkland	43.52	43.52

Table 1. Amount of land-use transitions observed in the historical maps from 1985 to 2010. e.g. 14.47% is the percentage of agriculture increase in 2010 from the existing area of agriculture in 1985 and 6.23%, 6.41%, 1.83% are the contributing portions to this increase from each land-use transition to agriculture

2.1 Model implementation

The CA model was written in IDL version 6.3 from ITT Visual Information Solutions (ITTVIS, 2007). IDL is an array-oriented interpreted language based on optimized C routines. As a consequence, an operation on an array can be performed at a speed unreachable by a traditional for-loop going through each element of an array. IDL also offers the advantages of being a multiplatform language, of having internal functions dealing with spatial data, and of being linked to ENVI, a remote sensing image analysis software.

The model implementation includes three main steps: 1) the definition of the cell size, neighborhood configuration, and driving factors, 2) the transition rule extraction and the model calibration, and 3) the simulation procedure.

2.1.1 Cell size, neighborhood configuration, and driving factors selection

Several studies have shown that the cell size and neighborhood configuration have an impact on the outcomes of raster-based CA models and should not be arbitrarily chosen (Chen and Mynett, 2003; Kocabas and Dragicevic, 2006; Ménard and Marceau, 2005; Moreno *et al.*, 2009; Pan *et al.*, 2010; Samat, 2006; Benenson, 2007). To guide the selection of the cell size, an examination of the historical land-use maps was done, which revealed that most land-use changes were occurring over four or more contiguous pixels. To reduce computation time while maintaining the desired level of spatial details for the study, the land-use maps were resampled at the resolution of 60 and 100 m using the nearest neighbor algorithm available in ArcGIS 9.1 (ESRI, 2005).

The neighborhood was designed to approximate a circle around a center cell. This decision was made in order to reduce spatial distortions, when compared to an extended Moore neighborhood, as every cell located at a given distance from the center cell is considered in the neighborhood (Li and Yeh, 2002b). The modeler can choose the desired number and size of concentric neighborhood rings around a cell. The different rings are all exclusive; a cell can only be located in a single ring, and there is no gap between two rings (Figure 3). Within each ring, the influence of the neighboring cells on the central cell is constant but this influence is different between rings. Consequently, the continuous distance function used in most CA models to represent the influence of neighborhood cells has been replaced by a discrete distance function. This approach has the main advantage of greatly simplifying the definition of the cells' influence as there is only one influence per ring. Moreover, these influences are dynamically found in the historical data and are not hard coded in the model, which allows the use of historical data at a different scale without changing the model.

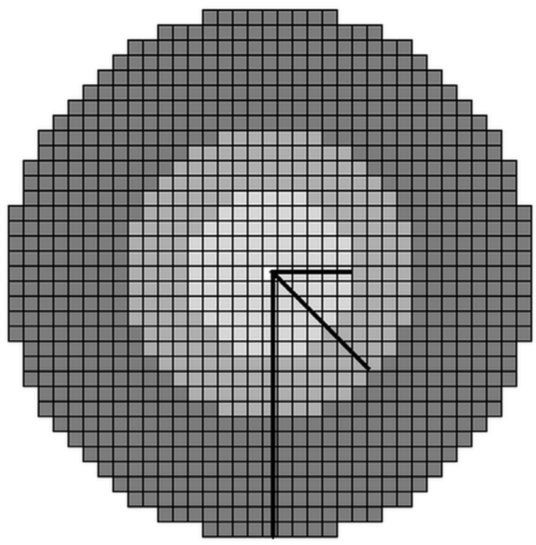


Fig. 3. Illustration of the neighborhood configuration used in the study corresponding to rings of 5, 9 and 17 cells

While testing all the possible combinations of cell size (60 m and 100 m) and neighborhood configuration was beyond the scope of this study, several combinations were tested to identify which ones provide the best simulation outcomes. Details regarding the sensitivity analysis that was conducted are provided in Section 2.1.3.

Land-use changes are complex spatial processes resulting from the interactions of socio-economic (e.g., population growth), biophysical (e.g., slope and soil quality), and geographic (e.g., proximity and accessibility to services) factors operating at different spatial and temporal scales (Liu and Phinn, 2003; Verburg *et al.*, 2004). In this study, in addition to the influence of the cells located within local and extended neighborhoods as previously described, four external factors were considered as parameters in the transition rules, namely the distance to Calgary city center, the distance to a main road, the distance to a main river, and the ground slope. Such factors are commonly quoted in the literature as influencing land-use changes (Fang *et al.*, 2005; Li and Yeh, 2002b; Pijanowski *et al.*, 2002; Wu, 2002). The aforementioned distances were calculated for each cell and each historical year using the Euclidian distance tool available in ArcGIS 9.1 (ESRI, 2006). The resulting distance files were stored as raster images of the same resolution and extent as the land-use maps.

2.1.2 Rule extraction and model calibration

The transition rule extraction and the model calibration include the following steps (Figure 4). First, the set of historical land-use maps along with the maps corresponding to the driving factors are read and the number of cells of a particular state in the neighborhood of each central cell is computed. For each type of land-use change, all the cells that have changed state in the historical land-use maps are identified. Frequency histograms are built to display the percentage of cells that have changed from one state to another when considering a particular driving factor and the cell state in the neighborhood. This provides quantitative information regarding the importance of each driving factor and neighborhood composition (i.e. state of the cells within the neighborhood) as being related to historical land-use changes within the study area. These histograms are interpreted by the modeler who identifies the ranges of values of each driving factor and neighborhood composition to be included in the conditional transition rules. This information is then automatically translated into mathematical transition rules.

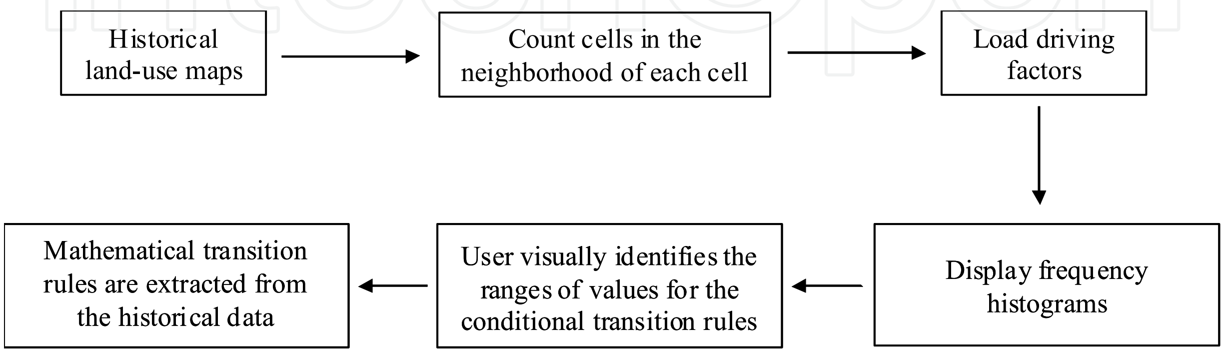


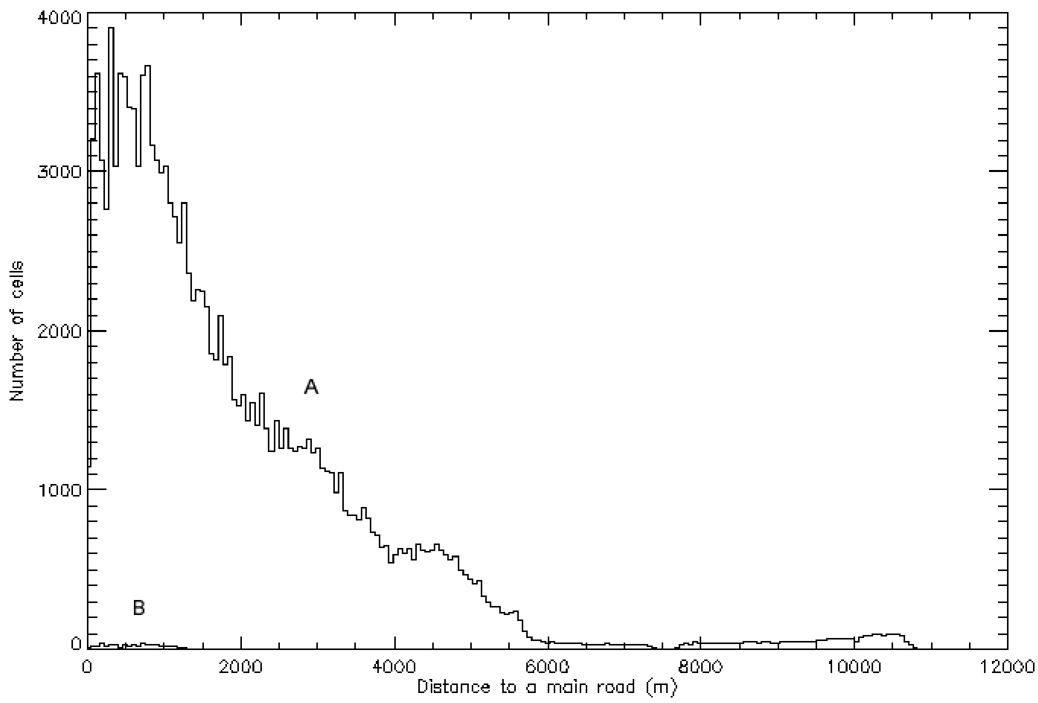
Fig. 4. Procedure for the extraction of the transition rules

Figure 5 provides an example of such a frequency histogram. The total number of Evergreen cells in the study area compared to the number of Evergreen cells that have changed to Built-up areas is first displayed to show the relative contribution of the later in the study area (Figure 5a). A detailed representation and analysis of the proportion of cells that have changed from Evergreen to Built-up areas when considering their distance to a main road (Figure 5b) reveals that 8% of these cells were located between 150 and 180 m of a main road while 98% of the cells were within 1250 m of a main road. At 1250 m, there is an inflexion point on the cumulative occurrence curve, expressing that this distance is critical for interpreting the influence of a main road on this land-use change. The further a cell was located from a main road, the less often it changed from Evergreen to Built-up area.

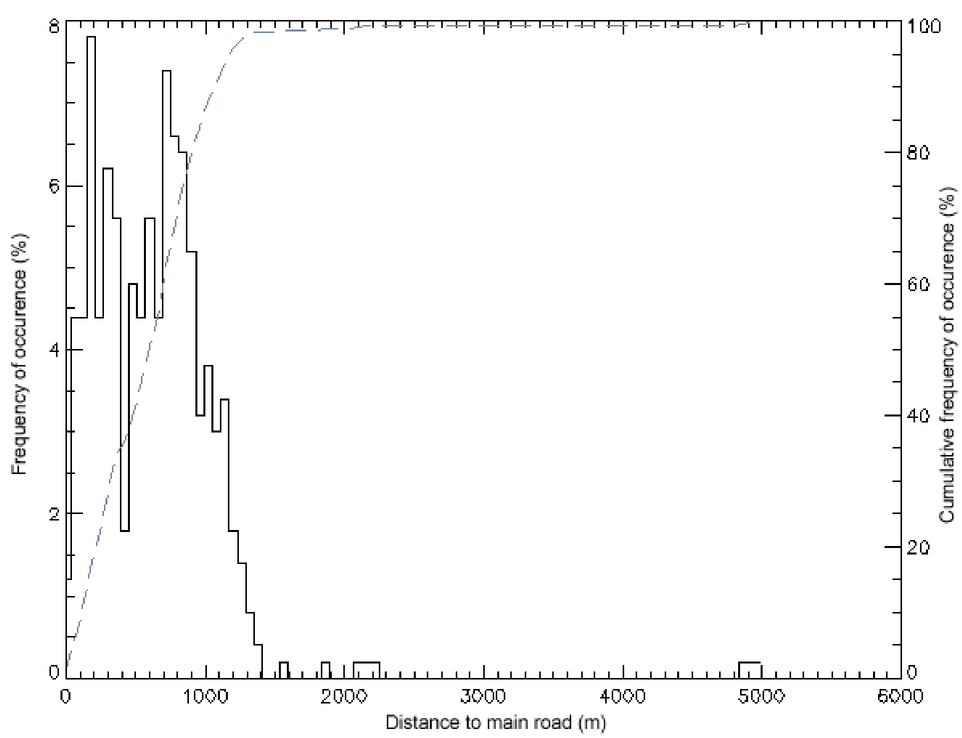
A graphical interface was designed to facilitate the interpretation of the frequency histograms and to allow a modeler to interactively select the ranges of values to be used for defining the conditional transition rules of the CA model (Figure 6). Each histogram can be displayed, allowing the modeler to change the bin size and to zoom in and out. By clicking on the histogram, the modeler identifies the ranges of values (minimum and maximum) for each neighborhood configuration, driving factor and cell state within that neighborhood. These values are stored in a table (Table 2) and further used to determine the conditional transition rules. An example of such a rule defined from Table 2 is:

If distance to a main road is between 0 and 427 m
and number of evergreen cells within the first neighborhood ring is between 0 and 17
and number of built-up cells within the second neighborhood ring is between 0 and 14
and number of agriculture cells within the third neighborhood ring is between 0 and 168
then the central Evergreen cell might change from Evergreen to Built-up area.

All possible transition rules are created by combining the identified ranges of values from the histograms.



a)



b)

Fig. 5. a) Frequency histogram comparing the total number of Evergreen cells located at a certain distance from a main road (A), with the number of Evergreen cells that have changed from Evergreen to Built-up areas when considering their distance to a main road (B); b) Frequency histogram displaying the percentage of cells that have changed from Evergreen to Built-up areas when considering their distance to a main road; the dashed curve represents the cumulative occurrence of the cells located at a certain distance from a main road

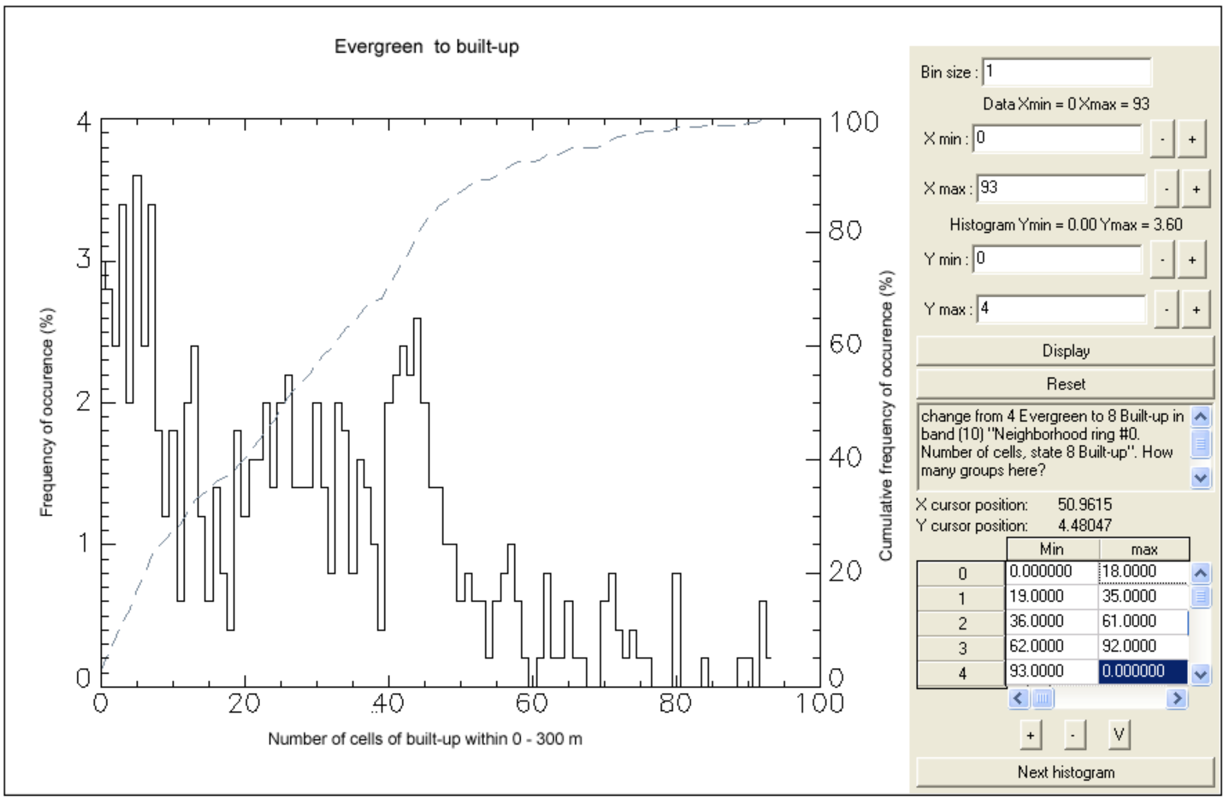


Fig. 6. Frequency histogram displaying the percentage of cells that have changed from Evergreen to Built-up when considering the number of Built-up cells within 300 m of these cells and graphical interface designed for the selection of the range of values to be considered in the conditional transition rules

Cell state	Distance to a main road (m)	Number of Evergreen cells located within the first neighborhood ring [0 to 300) m	Number of Built-up cells located within the second neighborhood ring [300 to 540) m	Number of Agriculture cells located within the third neighborhood ring [540 to 1020) m	…..
Evergreen	0 to 427	0 to 17	0 to 14	0 to 168	
	428 to 1408	18 to 50	15 to 59	169 to 258	
		51 to 74	60 to 92	259 to 377	

Table 2. Ranges of values identified from the frequency histogram to be used for determining the conditional transition rules

To convert the conditional rules into mathematical rules, the mean and standard deviation of the previously defined ranges of values are computed. These values become the coefficients of the parameters of the mathematical transition rules. In this model, the coefficients of each transition rule do not lead to a probability of change, but rather to a Resemblance Index (RI) that quantitatively describes the similarity between the neighborhood content of a cell at the time of the simulation and the neighborhood contents that have been used to generate the values of the parameters of the transition rule. If they are very similar, it is likely that the cell should change state for the corresponding type of land use. RI is inspired by the Minimum Distance to Class Mean remote sensing image classification algorithm (Richards, 2006). This algorithm calculates the mean point in the parameter space for pixels of known classes and then assigns unknown pixels to the class that is arithmetically closest. It is computed for every transition rule using Equation 1.

$$RI = \sum_{i=1}^{m} \frac{\left| n_i - \bar{x}_i \right|}{\sigma_i} \tag{1}$$

where m is the number of layers (corresponding to the number of driving factors plus the number of land-use classes multiplied by the number of neighborhood rings), n_i is the value in layer i, $\bar{x}_i$ is the mean value for layer i in the transition rule and σ_i is the standard deviation for layer i in the transition rule. If the standard deviation is zero for layer i, then $\frac{\left| n_i - \bar{x}_i \right|}{\sigma_i} = 0$ if $n_i = \bar{x}_i$ or otherwise equals positive infinity. Accordingly, $RI \in \Re^+$ and the smaller RI is, the more similar is the cell neighborhood configuration to the ones used to define the transition rule. The mathematical rules offer greater flexibility compared to the conditional rules as they reflect significant values for each type of land-use change and are more adaptable to the neighborhood composition than the conditional rules identified from specific observations in the historical dataset.

Table 3 presents some values representing the coefficients of the conditional and mathematical rules, respectively for three neighborhood configurations. The Min and Max columns are associated to the conditional transition rule, while the Mean and Standard deviation columns are related to the mathematical transition rule. An example of a mathematical rule defined from these values is,

$$\begin{aligned}
RI(\text{rule1, Evergreen to Agriculture}) &= \frac{\left|\text{D. main road} - 259.38\right|}{173.05} + \frac{\left|\text{D. city center} - 6\,272.57\right|}{1\,568.91} + \\
&\frac{\left|\text{D. river} - 3\,465.61\right|}{310.77} + \frac{\left|\text{Ground slope} - 3.23\right|}{1.79} + \frac{\left|\text{N0_Water} - 0.17\right|}{0.44} + \frac{\left|\text{N0_Evergreen} - 10.71\right|}{5.25} + \\
&\frac{\left|\text{N0_Deciduous} - 3.13\right|}{3.42} + \frac{\left|\text{N0_Agriculture} - 81.84\right|}{5.75} + \frac{\left|\text{N0_Rangeland / Parkland} - 0.04\right|}{0.29} + \\
&\frac{\left|\text{N0_Built - up} - 0.08\right|}{0.35} + \left\{0 \text{ if N0_Clear - cut} = 0; \infty \text{ otherwise}\right\} + \frac{\left|\text{N1_Water} - 0.44\right|}{0.86} + \\
&\frac{\left|\text{N1_Evergreen} - 19.24\right|}{10.0} + \frac{\left|\text{N1_Deciduous} - 4.75\right|}{3.9} + \frac{\left|\text{N1_Agriculture} - 170.93\right|}{11.51} + \\
&\frac{\left|\text{N1_Rangeland / Parkland} - 0.28\right|}{0.86} + \frac{\left|\text{N1_Built - up} - 0.33\right|}{0.63} + \left\{0 \text{ if N1_Clear - cut} = 0; \infty \text{ otherwise}\right\} +
\end{aligned}$$

$$\frac{|N2\_Water - 1.46|}{1.94} + \frac{|N2\_Evergreen - 86.53|}{43.49} + \frac{|N2\_Deciduous - 14.53|}{12.15} +$$

$$\frac{|N2\_Agriculture - 573.8|}{44.90} + \frac{|N2\_Rangeland\ /\ Parkland - 1.51|}{3.76} + \frac{|N2\_Built - up - 2.13|}{4.15} +$$

$$\frac{|N2\_Clear - cut - 0.02|}{0.14}$$

where Nx_LandUse is the number of cells of the corresponding land use within neighborhood rings of a cell. All the transition rules are stored in a file, so multiple simulations can be performed without re-calibrating the model.

Neighborhood ring	Layer	Min	Max	Mean	StdDev
Cell attributes	Cell state	Evergreen			
	Dist. to main road	0.0	780	259.38	173.05
	Dist. to city center	2473.86	8696.48	6272.57	1568.91
	Dist. to river	2979.53	4048.11	3465.61	310.77
	Ground Slope	0.0	7.29	3.23	1.79
[0-300) m	Nb cells state Water	0.0	2.0	0.18	0.44
	Nb cells state Evergreen	0.0	19.0	10.71	5.25
	Nb cells state Deciduous	0.00	11.00	3.13	3.43
	Nb cells state Agriculture	66.00	93.00	81.84	5.76
	Nb cells state Rangeland/Parkland	0.00	2.00	0.04	0.30
	Nb cells state Built-up	0.00	2.00	0.09	0.36
	Nb cells state Clear-cut	0.00	0.00	0.00	0.00
[300-540) m	Nb cells state Water	0.00	3.00	0.44	0.87
	Nb cells state Evergreen	0.00	32.00	19.24	10.00
	Nb cells state Deciduous	0.00	16.00	4.76	3.91
	Nb cells state Agriculture	151.00	192.00	170.93	11.52
	Nb cells state Rangeland/Parkland	0.00	4.00	0.29	0.87
	Nb cells state Built-up	0.00	3.00	0.33	0.64
	Nb cells state Clear-cut	0.00	0.00	0.00	0.00
[540-1020) m	Nb cells state Water	0.00	6.00	1.47	1.95
	Nb cells state Evergreen	2.00	205.00	86.53	43.49
	Nb cells state Deciduous	1.00	51.00	14.53	12.16
	Nb cells state Agriculture	430.00	656.00	573.80	44.91
	Nb cells state Rangeland/Parkland	0.00	14.00	1.51	3.76
	Nb cells state Built-up	0.00	24.00	2.13	4.15
	Nb cells state Clear-cut	0.00	1.00	0.02	0.15

Table 3. Example of data used to establish the conditional and mathematical transition rules

2.1.3 Simulation procedure

The simulation procedure includes these main steps.

- The mathematical transition rules previously defined and a land-use map corresponding to the beginning of the simulation are read.
- For each time step, the neighborhood configuration of every cell is read and the level of correspondence with the parameters of the transition rules is computed.
- With respect to the user-specified constraints and to the influence of each rule, the cells that change state do it according to the rule having the highest level of correspondence.
- To decide which cell should be associated to each type of land-use change, the model recursively sorts the type of land-use changes and for each of them selects the cell having the smallest RI value. Once the required number of cells associated to each type of land-use change is met or when no more cells can be assigned, the model writes the new land-use map and updates the statistics that correspond to the percentage of cells associated to each rule and each type of change.
- If the numbers of cells associated to each rule and each type of land-use change is different than the numbers found from the historical data and previous time steps, a correction is applied at the next time step. For example, if 200 cells are to change from Agriculture to Built-up area but only 150 of them can according to their neighborhood configuration, 50 additional cells will be set to change at the next time step.

Table 4 lists the land-use transitions that were considered during the simulations.

From	To
Evergreen	Agriculture
Deciduous	Agriculture
Evergreen	Built-up
Deciduous	Built-up
Agriculture	Built-up
Rangeland/Parkland	Built-up
Rangeland/Parkland	Agriculture
Agriculture	Rangeland/Parkland

Table 4. Land-use transitions considered during the simulations

Two sets of simulations were run. The first set was to test the model under various conditions. A sensitivity analysis to the cell size and neighborhood configuration was first carried out, followed by a sensitivity of the model to different ranges of values selected from the frequency histograms. Several ways of grouping the data values extracted from the histograms were tested for the calibration of the CA model: 1) the most dominant ranges of values ignoring flat areas, 2) the most dominant range of values concentrated around the mode, 3) the most dominant range of values dispersed away from the mode, and 4) the whole range of values of the histogram. Finally, to assess the importance of the selection of the external driving factors, simulations were conducted using four factors and different combinations of only three external factors.

In each case, the CA was calibrated using the land-use maps of 1985 to 2001, simulations were performed from 2001 to 2006, and the simulated map of 2006 was compared to the reference map of the same year using two similarity measures based on Kappa coefficients. The first measure is the standard Kappa coefficient, which expresses the percentage of agreement between two maps including both quantity and location information (Hagen, 2003; Pan *et al*., 2010; Visser and de Nijs, 2006). Three statistics were calculated, respectively referred to as Kappa, Kloc and Khisto. Khisto measures the quantitative similarity between two compared maps, while Kloc measures the similarity of the spatial allocation of categories between the maps. Kappa represents the general level of spatial agreement between two maps and is the product of Kloc and Khisto.

A drawback of the standard Kappa statistics is that they tend to over-estimate the agreement between a simulated map and a reference map because they do not take into account the percentage of cells that do not change state during the simulation period. In addition, they rely on a stochastic model of random allocation based on the sizes of the classes being compared to express the expected agreement. When simulating with a CA model, land-use allocation is not totally random since it depends on the initial conditions of the simulation. To compensate for these limitations, Van Vliet *et al*. (2010) introduced a coefficient of agreement called Kappa simulation that applies a more appropriate stochastic model of random allocation of class transitions that takes into account the information contained in the initial land-use map and the proportion of cells that does not change state over the simulation period. Three statistics were again calculated: Ksimulation that expresses the agreement between the simulated land-use map and the reference map, Ktransition that captures the agreement in terms of quantity of land-use transitions, and Ktransloc that measures the agreement between the two maps in terms of location of transition. Values of these coefficients vary from -1 to 1, the former value indicating a perfect disagreement between the two maps compared while the later indicating a perfect agreement. The standard and Kappa simulation coefficients were calculated using the Map Comparison Kit developed by the Research Institute for Knowledge System (RIKS BV, 2010).

To carry out a validation test, a simulation was conducted using the best combination of conditions described above from 2001 to 2006 and to 2010. A comparison was performed between the simulated maps and the reference maps for the years 2006 and 2010. An additional simulation was conducted from 1985 to 2010 to illustrate how the simulated land uses change over the whole period of time compared to the changes observed in the reference maps.

In all these simulations, a local constraint was applied to forbid built-up cells within the Tsuu T'ina nation. For the validation test where simulations were conducted from 2001 to 2010 and from 1985 to 2010, and where the selection of external driving factors was tested, a global constraint was also applied to restrict the number of built-up cells at each iteration based on an average estimated from the historical population trends.

3. Results

3.1 Sensitivity analyses

Table 5 presents the coefficients of agreement obtained when using a cell size of 60 m and 100 m, respectively. As expected, the values of the standard Kappa statistics tend to be high. They are also very similar and do not allow a discrimination among the results. However, the values of Ksim, Ktransloc and Ktransition all reveal that the simulation results obtained with 60 m are in higher agreement with the reference map than the results achieved using a cell size of 100 m.

The Ksim coefficient also shows that the choice of neighborhood configuration affects the simulation outcomes. Using only two rings in the neighborhood definition considerably reduces the performance of the model, while the best outcome is achieved when using three rings of respectively 5, 9 and 17 cells. This indicates that an extended neighborhood that covers a distance up to 1020 m is more appropriate in this study area to capture the zone of influence on central cells.

Cell size	Standard Kappa	Kloc	Khisto	Ksim	Ktransloc	Ktransition
60 m	0.853	0.875	0.975	0.047	0.085	0.551
100 m	0.850	0.873	0.974	0.043	0.078	0.546

Table 5. Kappa coefficients of agreement obtained when using a cell size of 60 m and 100 m

Neighborhood Configuration	Standard Kappa	Kappa simulation
3-5	0.845	0.015
3-5-15	0.850	0.037
5-9-14	0.852	0.044
5-9-15	0.852	0.045
5-9-16	0.853	0.046
5-9-17	0.853	0.047
5-12-17	0.852	0.045
6-9-15	0.849	0.031
6-14-18	0.852	0.043
6-14-19	0.852	0.045
7-10-15	0.852	0.042
7-10-16	0.852	0.044
7-10-17	0.852	0.044
7-13-17	0.850	0.034
7-14-18	0.852	0.043
7-14-19	0.852	0.044
7-14-20	0.852	0.045
7-15-19	0.852	0.043
8-12	0.847	0.024
8-15-19	0.852	0.042

Table 6. Kappa coefficients of agreement obtained when using different neighborhood configurations

The way ranges of values are selected from the frequency histograms to build the transition rules also affects the simulation outcomes (Table 7). The best results are achieved when the values are concentrated around the mode (Ksim = 0.069) compared to progressively more dispersed ranges of values (Ksim = 0.047 and 0.045). The worse result is achieved when a single range of values covering the whole histogram is selected (Ksim = 0.041).

Selection of parameter values	Kappa	KLocation	KHisto	Kappa simulation	KTransLoc	KTransition
Most dominant ranges of values	0.853	0.875	0.975	0.047	0.085	0.551
Values dispersed from the mode	0.853	0.875	0.975	0.045	0.081	0.551
Values concentrated around the mode	0.857	0.879	0.975	0.069	0.126	0.551
One group of values	0.852	0.874	0.975	0.041	0.074	0.551

Table 7. Kappa coefficients of agreement obtained when using different grouping of values from the frequency histograms for the definition of the transition rules

Simulation outcomes are also influenced by the number and selection of external driving factors (Table 8). Using four factors generates the highest agreement with the reference map both in terms of overall agreement (0.058) and location (0.140). The best combination of three factors includes distance to main road, distance to city center and distance to river, which were expected to play a major role in the increase of built-up areas. Ground slope also appears to be an important factor as revealed by the coefficients of agreement that are slightly lower than the ones obtained with the previous three factors. It can be observed that the amount of land-use transitions remains the same with the different combinations of factors; however, their spatial distribution changes as indicated by Ktransloc.

Factor selection	Standard Kappa	Kloc	Khisto	Ksim	KtransLoc	Ktrans
Dist. to main road Dist. to city center Dist. to river Ground Slope	0.866	0.876	0.989	0.058	0.140	0.411
Dist. to main road Dist. to city center Dist. to river	0.864	0.874	0.989	0.042	0.102	0.411
Dist. to main road Dist. to city center Ground Slope	0.864	0.873	0.989	0.038	0.094	0.411
Dist. to main road Dist. to river Ground Slope	0.864	0.874	0.989	0.041	0.100	0.411
Dist. to city center Dist. to river Ground Slope	0.864	0.874	0.989	0.041	0.101	0.411

Table 8. Kappa coefficients of agreement obtained when using four external driving factors compared to the combinations of only three factors

3.2 Results obtained with the best combination of conditions

The Kappa coefficients of agreement obtained when running simulations from 2001 to 2006 and to 2010 using the best combination of conditions are presented in Table 10. The agreement is higher for the year 2006 compared to the year 2010. A more detailed analysis of the results provided by the per-class Kappa simulation coefficients for the years 2006 and 2010 indicates that in terms of number of transition, the model achieves a relatively good agreement with the reference maps (values between 0.371 and 0.541), except for the class built-up where the values are slightly over 0.2 (Tables 11 and 12). The values obtained for Ktransloc are lower than those obtained for Ktransition indicating that the model is better at allocating the right amount of transition rather than their location.

	Year	
	2006	**2010**
Standard Kappa	0.869	0.782
Kappa simulation	0.075	0.057

Table 10. Overall Kappa coefficients of agreement obtained when running simulation from 2001 to 2010 and comparing the results with the reference maps of 2006 and 2010

	Built-up	**Rangeland/ parkland**	**Agriculture**	**Deciduous**	**Evergreen**
Standard Kappa	0.816	0.669	0.860	0.906	0.947
Kloc	0.817	0.684	0.865	0.917	0.948
Khisto	0.999	0.978	0.994	0.988	0.999
Kappa simulation	0.009	0.140	0.096	0.084	0.065
KtransLoc	0.045	0.259	0.178	0.165	0.174
Ktransition	0.211	0.541	0.536	0.508	0.376

Table 11. Per-class Kappa coefficients of agreement obtained when running simulation from 2001 to 2006 and comparing the simulated map of 2006 with the reference map of 2006

	Built-up	**Rangeland/ parkland**	**Agriculture**	**Deciduous**	**Evergreen**
Standard Kappa	0.706	0.537	0.779	0.822	0.905
Kloc	0.793	0.553	0.806	0.834	0.930
Khisto	0.890	0.971	0.967	0.986	0.973
Kappa simulation	0.011	0.095	0.078	0.079	0.046
KtransLoc	0.054	0.198	0.148	0.171	0.123
Ktransition	0.204	0.483	0.530	0.460	0.371

Table 12. Per-class Kappa coefficients of agreement obtained when running simulation from 2001 to 2010 and comparing the simulated map of 2010 with the reference map of 2010

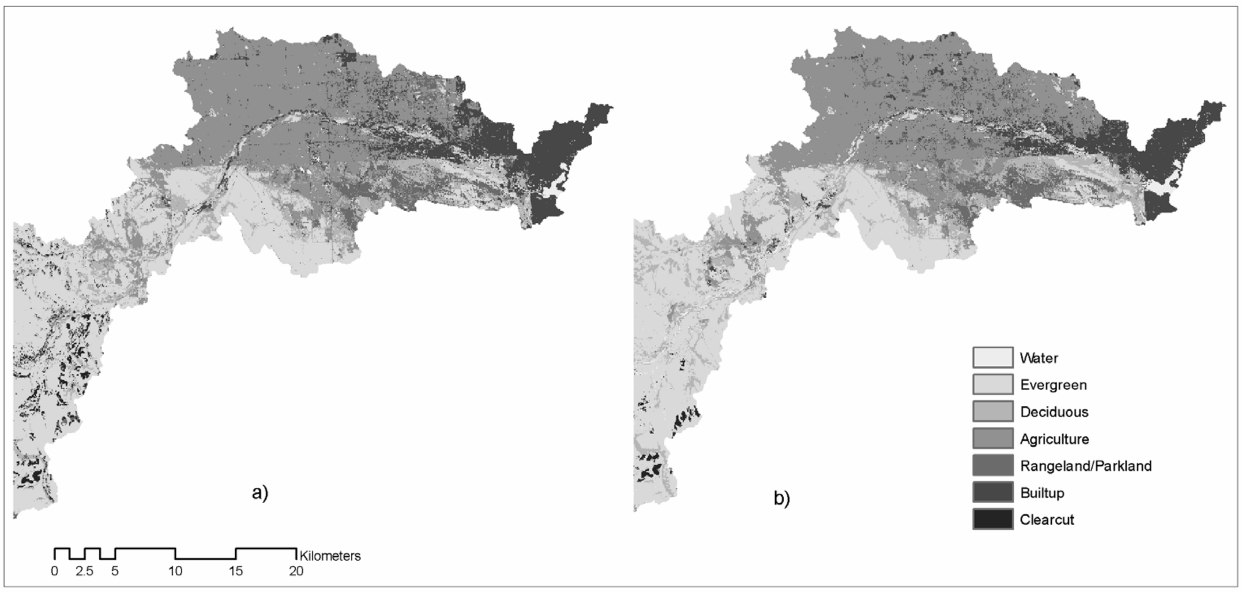


Fig. 7. Comparison of the reference map of 2010 (a) with the simulated map of the same year (b)

A visual comparison of the 2010 reference and simulated maps shows an under-estimation of built-up cells by the model. This is mainly due to the considerable urban growth that occurred during the period 2006-2010 that is not accounted for in the global constraint of the model. The simulation of agricultural areas is affected by the above since most built-up cells in the reference map are located within the agricultural areas.
Simulations run from 1985 to 2010 provide additional details and reveal that the CA model under-estimates the quantity of built-up areas (difference of 4.37% in 2010) and over-estimates agriculture (difference of 5.73% in 2010) (Table 9). The proportion of rangeland/parkland simulated by the model does not differ greatly from the proportion observed in the reference maps. The under-estimation of deciduous can be explained by the fact that the CA model does not simulate the transition to this class.

Land-use	1985	1992	1996	2001	2006	2010
Evergreen	36.87	35.80(-0.47)	35.47(-0.40)	35.15(0.19)	34.61(0.18)	34.19(1.34)
Deciduous	14.08	12.67(-1.50)	12.19(-1.85)	11.68(-2.45)	10.93(-2.81)	10.33(-2.29)
Agriculture	36.16	37.56(1.29)	38.12(2.03)	38.65(3.91)	39.43(3.82)	40.06(5.73)
Rangeland /Parkland	4.48	5.07(0.75)	5.04(0.77)	4.99(0.62)	5.14(0.75)	5.26(0.25)
Built-up	6.12	6.61(-0.80)	6.90(-1.31)	7.25(-2.19)	7.60(-2.18)	7.88(-4.37)

Table 9. Percentage of the study area covered by the main land uses in the simulation results from 1985 to 2010. The variation with the original land-use maps is shown in parentheses; a positive value indicates an over-estimation by the model while a negative value indicates the opposite.

The fact that values of Kappa simulation and Ktransloc are sometimes relatively low might be explained by the large number of land-use transitions considered in this study and the difficulty of capturing the dynamics specific to each type of transition. To obtain a preliminary assessment of how the model would perform with a reduced number of land-use classes and transitions, a simulation was run with aggregated land-use maps in which the number of land-use classes was reduced to five (water, forest, agriculture, built-up, and Tsuu T'ina nation) and only four transitions were considered, namely forest to agriculture, forest to built-up, agriculture to forest, and agriculture to built-up. The model was calibrated over the period 1985-2001 and the simulation was run from 1985 to 2006 using three external driving factors (distance to Calgary city center, distance to a main road, distance to the main river). Higher values were obtained for the three Kappa simulation statistics calculated (Table 14), confirming that the simulation results could be improved by either reducing the number of land-use transitions in the model or improving the rules for some of the land-use transitions considered.

	1992	1996	2001	2006
Kappa	0.947	0.941	0.929	0.917
KLocation	0.951	0.947	0.938	0.939
KHisto	0.996	0.994	0.991	0.976
Kappa Simulation	0.304	0.262	0.251	0.227
KTransLoc	0.429	0.377	0.336	0.341
KTransition	0.709	0.695	0.747	0.665

Table 14. Overall Kappa coefficients of agreement obtained when running simulation with a reduced number of land-use transitions.

4. Conclusion

While the potential of CA models is increasingly acknowledged for land-use change studies, their calibration and the evaluation of their performance remains a challenge. In this research, we developed a calibration method that allows the modeler to interactively obtain information about historical land-use changes and the factors associated to these changes in order to automatically derive conditional and mathematical transition rules. When testing the applicability of this model in a study area in southern Alberta, sensitivity analyses were conducted to evaluate the influence of various conditions involved in the calibration of the model, including the cell size, the neighborhood configuration, the selection of the parameter values, and the number of driving factors. These analyses indicate that the simulation outcomes are affected by the selection of these conditions and that there exist no method to *a priori* identify the most adequate combination.

Sensitivity of raster-based CA models to cell size and neighborhood configuration has been recognized by several authors over the last years. One approach to overcome this sensitivity to scale is the implementation of object-based CA models with the inclusion of a dynamic neighborhood as proposed by Moreno *et al.* (2009, 2010) and others (Hamman *et al.*, 2007). While such models are computationally intensive and the handling of the topology cumbersome, they appear as a promising approach to better capture the meaningful entities composing a landscape along with their evolution.

The calibration technique described in this paper provides useful insights regarding the number and choice of external driving factors that should be considered in the calibration. When taking into account external factors in addition to the influence of the cells within extended neighborhoods, the number of possible combinations of factors becomes too high to be thoroughly evaluated by a simple sensitivity analysis. Other approaches based on data mining techniques might be useful in this context to guide the selection of driving factors for the calibration of the model (Wang *et al*., in press).

Kappa simulation is a recently proposed coefficient of agreement specifically adapted to the context of evaluating the performance of a CA model (Van Vliet *et al*., 2010). While additional studies are needed to fully assess the interpretation potential of this coefficient, it appears very useful in this study to capture differences in simulation results when the standard Kappa was not sensitive enough to provide discrimination. In particular, it indicates that the CA model generates a relatively high agreement in terms of amount of land-use transitions, while the agreement in terms of location is lower. The fact that the values of the coefficients increase when reducing the number of land-use transitions considered in the model also reveals that additional external driving factors might be necessary to fully capture the dynamics of the study area.

When interpreting the values of these coefficients however, we must keep in mind that they inform on the agreement between two maps on a cell-by-cell basis, without considering a slight displacement that might occur among the cells being compared. In addition, the comparison is performed between two possible states of the area being studied, respectively generated by the model and from observations acquired at a specific moment in time. These two states might differ, which does not necessarily imply that the simulated outcomes are 'wrong'.

5. Acknowledgments

This project was initiated in collaboration with the Calgary Regional Partnership (CRP) who provided financial support to J.-G. Hasbani. We thank Cheng Zhang from the University of Calgary for producing the historical land-use maps used in the project and Jasper Van Vliet for the stimulating discussions regarding Kappa simulation. Additional funding was provided by a NSERC Discovery grant awarded to D. Marceau and by GEOIDE, the Canadian Network of Centers of Excellence in Geomatics.

6. References

Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S., 2008. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science* 22(9): 943-963.

Benenson, I., & Torrens, P., 2004. *Geosimulation: Automata-based modeling of urban phenomena*, Wiley and Sons.

Benenson, I., 2007. Warning! The scale of land-use CA is changing! *Computers, Environment and Urban Systems* 31(2): 107-113.

Bone, C., Dragicevic, S., & Roberts, A., 2006. A fuzzy-constrained cellular automata model of forest insect infestations. *Ecological Modelling* 192: 107-125.

Calgary Economic Development, 2010, http://www.calgaryeconomicdevelopment.com/liveWorkPlay/Live/demographics.cfm

Chen, Q., & Mynett, A. E., 2003. Effects of cell size and configuration in cellular automata based prey-predator modelling. *Simulation Modelling Practice and Theory* 11(7-8): 609-625.

Clarke, K. C., Hoppen, S., & Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B* 24(2): 247-261.

Erlien, C. M., Tuttle, J. P., McCleary, A. L., Mena, C.F., & Walsh, S.J., 2006. Complexity theory and spatial simulations of land use/land cover dynamics: the use of what if scenarios for education, land management, and decision-making. *Geocarto International* 21(4): 67-74.

ESRI, 2006. ArcGIS 9.1 Users' manual. ESRI, Redlands, California.

Fang, S., Gertner, G. Z., Sun, Z., & Anderson, A.A., 2005. The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning* 73: 294-306.

Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science* 17(3): 235-249.

Hammam, Y., A. Moore, & Whigham, P. (2007). The dynamic geometry of geographical vector agents. *Computers, Environment and Urban Systems* 31: 502-519.

Jantz, C. A., & Goetz, S. J., 2005. Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science* 19(2): 217-241.

Jantz, C. A., Goetz, S. J., & Shelley, M.K., 2003. Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environment and Planning B* 30: 251-271.

Jenerette, D. G., & Wu, J., 2001. Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. *Landscape Ecology* 16:611-626.

Kocabas, V., & Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems* 30: 921-953.

Lau, K. H., & Kam, B. H., 2005. A cellular automata model for urban land-use simulation. *Environment and Planning B* 32: 247-263.

Li, X., & Yeh, A. G.-O., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science* 14(2): 131-152.

Li, X., & Yeh, A. G.-O., 2002a. Integration of principal components analysis and cellular automata for spatial decision making and urban simulation. *Science in China*, 45(6): 521-529.

Li, X., & Yeh, A. G.-O., 2002b. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science* 16(4): 323-343.

Li, X., & Yeh, A. G.-O., 2004. Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science* 18(8): 723-744.

Liu, Y., & Phinn, S. R., 2003. Modelling urban development with cellular automata incorporating fuzzy-set approaches. *Computers, Environment and Urban Systems* 27: 637-658.

Ménard, A., & Marceau, D. J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B* 32: 693-714.

Ménard, A., & Marceau, D. J., 2007. Simulating the impact of forest management scenarios in an agricultural landscape of southern Quebec, Canada, using a geographic cellular automaton. *Landscape and Urban Planning* 79(3-4): 253-265.

Moreno, N., Wang, F., & Marceau, D.J., 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems*, 33: 44-54.

Moreno, N., Wang, F., & D.J. Marceau, 2010. A geographic object-based approach in cellular automata modeling. *Photogrammetric Engineering and Remote Sensing* 76(2): 183-191.

Ohgai, A., Gohnai, Y., & Watanabe, K., 2007. Cellular automata modeling of fire spread in built-up areas: A tool to aid community-based planning for disaster mitigation. *Computers, Environment and Urban Systems* 31(4): 441-460.

Pan, Y., Roth, A., Yu, Z., & Doluschitz, R., 2010. The impact of variation in scale on the behavior of a cellular automata used for land-use change modeling. *Computers, Environment and Urban Systems* 34: 400-408.

Pijanowski, B. C., Brown, D.G., Shellito, B.A., & Manik, G.A. 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, Environment and Urban Systems* 26: 553-575.

Richards, J. A., 2006. *Remote sensing digital image analysis: An introduction*, Springer, 439 p.

Research Institute for Knowledge System (RIKS BV), 2010, http://www.riks.nl/mck

Samat, N., 2006. Characterizing the scale sensitivity of the cellular automata simulated urban growth, A case study of the Seberang Perai Region, Penang State, Malaysia Computers, *Environment and Urban Systems* 30(6): 905-920.

Santé, I., Garcia, A. M., Miranda, D., & Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape and Urban Planning* 96: 108-122.

Shan, J., Alkheder, S., & Wang, J., 2008. Genetic algorithms for the calibration of cellular automata urban growth modeling. *Photogrammetric Engineering and Remote Sensing* 74(10): 1267-1277.

Shen, Z., Kawakami, M., & Kawamura, I., 2009. Geosimulation model using geographic automata for simulating land-use patterns in urban partitions. *Environment and Planning B* 36: 802-823.

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

Straatman, B., White, R., & Engelen, G., 2004. Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems* 28: 149-170.

Sui, D. Z., & Zeng, H., 2001. Modeling the dynamics of landscape structure in Asia's emerging desakota regions, a case study in Shenzhen. *Landscape and Urban Planning* 53(1-4): 37-52.

Sun, T., & Wang, J., 2007. A traffic cellular automata model based on road network grids and its spatial and temporal resolution's influences on simulation. *Simulation Modelling Practice and Theory* 15: 864-878.

Van Vliet, J., White, R., & Dragicevic, S., 2009. Modeling urban growth using a variable grid cellular automata. *Computers, Environment and Urban Systems* 33: 35-43.

Van Vliet, J., Bregt, A. K., & Hagen-Zanker, A., 2010. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. Submitted to *Ecological Modelling*.

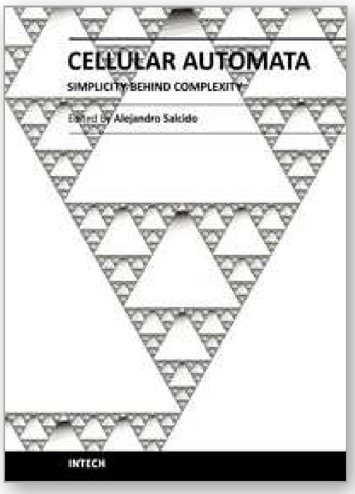
Verburg, P. H., de Nijs, T. C. M., Van Eck, J.R., Visser, H., & de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems* 28(6): 667-690.

Visser, H., & de Nijs, T., 2006. The Map Comparison Kit. *Environmental Modelling and Software* 21(3): 346-358.

Wang, F., Hasbani, J-G., Wang, X., & Marceau, D. J., 2011. Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set theory. *Computers, Environment and Urban Systems*, in press.

White, R., Engelen, G., & Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B* 25: 323-343.

Wu, F., 2002. Calibration of stochastic cellular automata, the application to rural-urban land conversions. *International Journal of Geographical Information Science* 16(8): 795-818.



Cellular Automata - Simplicity Behind Complexity
Edited by Dr. Alejandro Salcido

ISBN 978-953-307-230-2
Hard cover, 566 pages
Publisher InTech
Published online 11, April, 2011
Published in print edition April, 2011

Cellular automata make up a class of completely discrete dynamical systems, which have became a core subject in the sciences of complexity due to their conceptual simplicity, easiness of implementation for computer simulation, and their ability to exhibit a wide variety of amazingly complex behavior. The feature of simplicity behind complexity of cellular automata has attracted the researchers' attention from a wide range of divergent fields of study of science, which extend from the exact disciplines of mathematical physics up to the social ones, and beyond. Numerous complex systems containing many discrete elements with local interactions have been and are being conveniently modelled as cellular automata. In this book, the versatility of cellular automata as models for a wide diversity of complex systems is underlined through the study of a number of outstanding problems using these innovative techniques for modelling and simulation.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821