Environment and Planning B: Planning and Design 2008, volume 35, pages 647 – 665

doi:10.1068/b33093

VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects

Niandry Moreno ¶
Geocomputing Laboratory, University of Calgary, Calgary, Alberta T2N 1N4, Canada;
e-mail: morenos@ula.ve
André Ménard
Planning and Development Research Center (CRAD), Laval University, Quebec G1K 7P4,
Canada; e-mail: menarda@sympatico.ca
Danielle J Marceau §
Geocomputing Laboratory, Department of Geomatics Engineering, Schulich School of
Engineering, University of Calgary, 2500 University Drive NW Calgary, Alberta T2N 1N4,
Canada; e-mail: marceau@geomatics.ucalgary.ca
Received 17 August 2006; in revised form 25 May 2007;

Abstract. Cellular automata (CA) can reproduce global patterns and behavior from local interactions of cells and they are used increasingly to simulate complex natural and human systems. Among their attributes are their computational simplicity and their explicit representation of space and time. However, the classic definition of CA limits their application to problems that involve a discrete space, and similar rules and neighborhoods for all cells. In addition, the standard raster-based CA model is sensitive to spatial scale. This paper presents a new vector-based geographic cellular automata model, called the VecGCA model, which defines space as a collection of irregular geographic objects. Each object has a geometric representation (a polygon) that evolves through time according to a transition function that depends on the influence of neighboring polygons. In this model, the neighborhood is defined as the region of influence on each geographic object, and the neighbors are all geographic objects located within the region of influence. An innovative aspect of the VecGCA model is that the procedure allows geometric transformation of objects. The area of a polygon (representing an object) is reduced in the region that is nearest to the neighbor that exerts an influence on it, and the area of that neighbor is increased accordingly. The proposed model was tested with real data and compared with a raster-based CA model to simulate land-use changes in an agroforested area in southern Quebec, Canada. The model was validated using two land-use maps, produced from satellite Landsat Thematic Mapper imagery, which were acquired in 1999 and 2002. The results obtained show that VecGCA can represent well the dynamics in the study area through an adequate evolution of the geometry of the geographic objects which are independent of the cell size, whereas, to generate similar outcomes in the raster-based CA model, a sensitivity analysis must be conducted to determine which cell size is needed. The geometric transformation procedure introduced in the VecGCA model executes the change of shape of a geographic object by changing its state in a portion of its surface, allowing a more realistic representation of the evolution of the landscape.

1 Introduction

Cellular automata (CA) are dynamic systems defined by a large tessellation of finite-state cells whose states are updated at discrete time steps according to deterministic or probabilistic rules, which dictate how the state of each cell might change on the basis of the state of its neighboring cells. These systems are able to reproduce complex global behavior and patterns from simple local interactions of individual cells (Wolfram, 1984). Owing to their computational simplicity and their explicit representation of space and time, CA constitute a powerful tool to model complex natural and human systems.

¶ Also: Centro de Simulación y Modelos (CESIMO), Universidad de Los Andes, Mérida, Venezuela.
§ Corresponding author.

However, the assumption of regularity, uniformity, and homogeneity in the classic definition of CA makes their application difficult for simulating real-world phenomena.

To overcome these limitations, several CA extensions have been proposed. In a pioneer study, Couclelis (1985) presented a model allowing the separation of the neighborhood set and the transition rules for each cell. Takeyama and Couclelis (1997) developed a mathematical framework, called Geo-Algebra, to integrate CA and geographic information systems (GIS), which expresses the modeling paradigm of CA in the form of map equations. In addition, Geo-Algebra generalizes the structure of standard CA to accept arbitrary, spatially variant neighborhoods and transition rules. Other research conducted by White and Engelen (2000) and by O'Sullivan (2001b) extends the definition of neighborhood to the set of all cells (adjacent or not) that influence the state of a particular cell, referred to as a radius of influence and a graph of influence, respectively.

Other authors have focused their attention on different aspects related to CA modeling, including the calibration and validation of CA models (Dietzel and Clarke, 2006; Li and Yeh, 2001; 2004; Liu and Phinn, 2003; Straatman et al, 2004; Wu, 2002), the definition of neighborhood as directed graphs (O'Sullivan, 2001a; 2001b), the analysis of neighborhood relationships (Verburg et al, 2004), and the use of data integrated within a GIS to obtain attribute values and training data (Li and Yeh, 2000; 2002).

Thanks to all these advances, in the last few years numerous studies involving the application of CA models in a geographic context have been undertaken. Geographic CA have been used to simulate land-use changes and land-cover changes (de Almeida et al, 2003; Li and Yeh, 2002; Ménard and Marceau, 2007; Parker et al, 2003; White et al, 2000; Wu, 2002), fire propagation (Clarke et al, 1994; Favier et al, 2004), vegetal succession (Rietkerk et al, 2004; Thiéry et al, 1995), and urban growth and development (Benenson and Torrens, 2004; Clarke et al, 1997; Lau and Kam, 2005; Li and Yeh, 2000; Liu and Phinn, 2003; White and Engelen, 2000).

These applications have been achieved using a discrete space representation and a regular tessellation of cells of the same size and shape similar to the GIS raster model. However, recent studies have demonstrated that such raster-based CA models are sensitive to spatial scale. Chen and Mynett (2003) revealed that the choice of a particular cell size and neighborhood configuration has a clear effect on the resulting spatial patterns in their CA based prey–predator model. Jenerette and Wu (2001) tested two cell sizes in their CA model that was developed to simulate the urbanization process in Phoenix, Arizona. They obtained reliable results with the coarser resolution, but poor results with the finer resolution. Jantz and Goetz (2005) showed the influence of cell size on the SLEUTH urban-growth model; their results demonstrated that the model is able to capture reliably the rate of growth across different cell sizes, but differences in the ability to simulate growth patterns were substantial. Kocabas and Dragicevic (2006) analyzed the behavior of a GIS-based CA urban-growth model using sensitivity analysis, and demonstrated that CA-model responses are different depending on the spatial resolution and the neighborhood configuration. In an exhaustive study undertaken to assess the spatial-scale sensitivity in a land-use CA model, Ménard and Marceau (2005) revealed that the choice of cell size and neighborhood configuration has a considerable impact on the simulation results in terms of land-cover area and spatial patterns. They advocate the development of alternative CA models, such as vector-based or object-based models where space is defined as a collection of irregular polygons that correspond to real entities in the study area, to mitigate scale sensitivity.

Some researchers have begun to implement irregular space in CA models through the use of Voronoi polygons (Flache and Hegselmann, 2001; Shi and Pang, 2000), but this space definition does not necessarily correspond to real-world entities. A Voronoi polygon represents a region grouping together the set of points closest to a spatial object, but it does not represent the spatial object itself. In other studies, the GIS vector format is used to define space, where each polygon represents a real-world entity. The vector cellular automaton model presented by Hu and Li (2004) is an extended CA model based on geographic objects, where a geographic object is the conceptual representation of a real entity such as a city, a farm, a land parcel, or a school. Each geographic object has a spatial representation under the Cartesian coordinate system (a point, a line, a polyline, or a polygon). Neighborhood relationships among geographic objects are defined using Voronoi diagrams. The transition rules define the next state of a geographic object, determined by the state and area of its neighbors, the distance between the geographic object and its neighbors, and the length of the common border. This model presents two disadvantages. The first is the lack of an explicit definition of the neighborhood relationships because the Voronoi diagram is automatically generated from the vector map that represents space. The second disadvantage is that the model does not allow a change of shape or size of the objects, but only their change of state.

Recently, another irregular CA model designed as a tool for urban planning was presented by Stevens et al (2007), where space is defined as a collection of irregular cadastral land parcels and the neighbors are composed of all adjacent parcels, parcels accessible from a road, and parcels within a buffer zone. The appearance of new parcels is based on a set of predefined parcels that change state from undeveloped to developed. However, this model also does not allow the change of shape or size of the land parcels.

This inability to change size or shape constitutes an important limitation since such changes occur continuously in the real world and should be taken into account. For example, urbanization involves the expansion of urban areas at the expense of nondeveloped areas, and this produces geometric transformations in the urban and nonurban patches forming that geographic area. Similar transformations characterize deforestation or any other land-use and land-cover changes.

This study has been undertaken to achieve the following objectives:

1. To present a new vector-based CA model that overcomes the problem of cell-size sensitivity of classical raster-based CA models and of recent implementations of vector CA by allowing an irregular space tessellation where the neighborhood definition and the transition rules are connected to the real properties of each geographic object within the study area, and that allows geometric transformations of the geographic objects as a result of the transition functions.
2. To test the vector-based geographic cellular automata (VecGCA) model with real data to simulate land-use changes in an agroforested area in southern Quebec, Canada, and compare the results with those obtained using a classical raster-based CA model applied to the same area under the same conditions.

A detailed description of the proposed VecGCA model is presented, including the conceptual model and the implementation details. The methodology for the definition of the VecGCA model to simulate land-use changes and the comparison with a raster-based CA model is described. Finally, results, discussion and conclusions are presented.

2 The proposed VecGCA model

The proposed model is an extension of classical CA model, where the space, the neighborhood, and the transition rules allow the representation of an irregular complex dynamic space.

2.1 Conceptual model

The conceptual model defines the three components of VecGCA that correspond to modifications to the classical CA formalism and the procedure of geometric transformations that characterize the evolution of the geographic objects.

2.1.1 *Space definition*

A space is composed of a collection of georeferenced geographic objects of irregular shape that represent real-world entities. Every geographic object in the space is represented as a polygon which is in a specific state and can define its proper transition function and neighborhood. Each geographic object is connected to others through adjacent sides and together they compose the whole space of a study area.

2.1.2 *Neighborhood definition*

The neighborhood is a key component in the VecGCA model, because it defines which objects determine the change of shape of a central geographic object. In geographic applications, such as land-use – land-cover changes or urban development, a geographic entity is generally influenced by adjacent or nonadjacent entities separated by a distance *d*.

The VecGCA model defines the neighborhood as an external buffer (region of influence) around each geographic object and the neighbors are all the objects (adjacent or nonadjacent) that are totally or partially within the region of influence (see figure 1). The size of the buffer that delineates the region of influence is expressed in length units and is selected by the user.

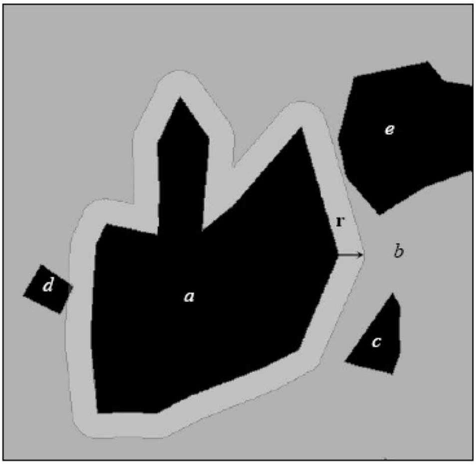


Figure 1. Neighborhood as defined in VecGCA (vector-based geographic cellular automata) model. For *a*, the neighborhood defined by **r** includes the objects *b* and *d* but not *e* or *c*.

In order to determine the neighbor priority during the execution of the transition function, an influence function is applied. This function quantifies the influence of each neighbor on a specific geographic object. Its value varies between 0 and 1, where 0 indicates no influence and 1 the highest influence. Three main parameters are considered in that function: the neighbor's area within the neighborhood, the probability that a geographic object changes its state to the state of its neighbor, and the distance between a neighbor and a geographic object [equation (1)]. In order to consider distance greater than zero and to give relevance to adjacent and non-adjacent neighbors, the distance between the centroids of the polygon is used.

Other parameters could be included to represent specific aspects of the dynamics of the study area.

$$g_{ab} = g[A(t)_a, P_{b\rightarrow a}, d_{ab}] \,, \tag{1}$$

where

g_{ab} is the influence of the neighbor *a* on the object *b*,

$A(t)_a$ is the area of the neighbor *a* within the neighborhood of the object *b* at time *t*,

$P_{b\rightarrow a}$ is the probability of *b* changing its state to that of *a*,

d_{ab} is the distance between the centroid of the neighbor *a* and the centroid of the object *b*.

2.1.3 *Definition of the transition function*

A transition function is defined for each geographic object to quantify its area that changes state. The area that changes is related to the area of the neighbor within the neighborhood and its influence on the specific geographic object [equation (2)]. The transition function has 0 as its lower limit, when the influence of the neighbor is smaller than a threshold value (λ), and the total area of the geographical object as its upper limit, when the whole area of the geographic object changes state. The threshold value represents the resistance of a geographic object to change its state to the state of its neighbor.

$$f_b(t+1) = f[A(t)_a, g_{ab}] \,, \tag{2}$$

where f_b is the transition function of object *b*.

The transition function is evaluated for each neighbor of the geographic object. The change of state of a portion or the totality of the geographic object is performed in the procedure of geometric transformations. This procedure is executed *n* times, once for each neighbor for which the transition function is greater than zero. The execution order is descendant from the neighbor having the highest influence to the neighbor having the lowest influence.

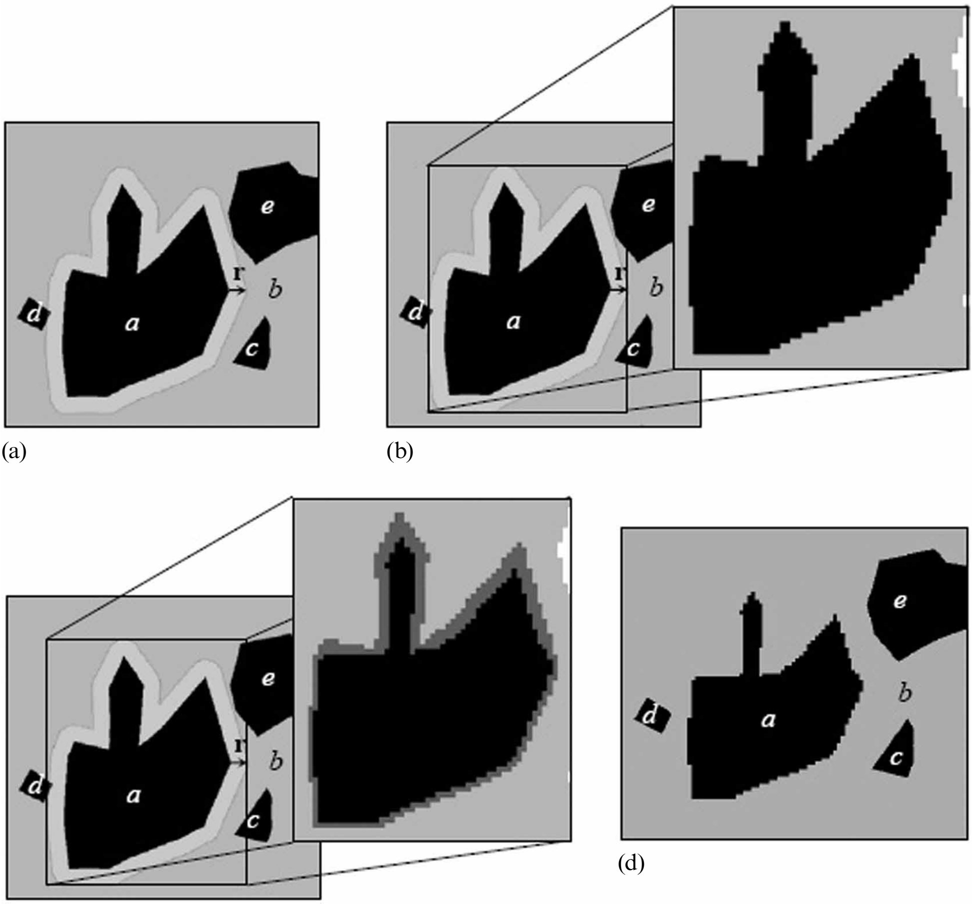
The geometric transformations are synchronous; all geographic objects forming a study area are considered during the same time step and their changes of shape are executed as needed.

2.1.4 *Procedure of geometric transformations*

The procedure of geometric transformations reduces the area of the geographic object by removing a quantity (expressed in area unit) from the region nearest to the corresponding neighbor. The principle is to remove small portions of the geographic object at the frontier with its neighbor until the area calculated in the transition function is removed. The problem is to determine the shape and size of these small portions to be removed from the geographic object. The solution applied here is to rasterize the geographic object using a regular grid, the resolution of which is defined by the user, and eliminate the necessary cells (nearest to each respective neighbor) to satisfy the area that has been calculated by the transition function. The number of cells to be eliminated is given by equation (3).

$$N_c = \frac{f_b(t+1)}{R^2} \,, \tag{3}$$

where N_c is the number of cells and *R* is the resolution of rasterization. The removed cells define a new object that is later combined with the corresponding neighbor. The cells that remain in the state of the geographic object are vectorized and define the new geometric representation of the geographic object.



(a) (b) (c) (d)

Figure 2. Procedure of geometric transformation of polygons.

A hypothetic example illustrating this procedure is presented in figure 2. In figure 2(a), the polygon *b* is a neighbor of the object *a* and exerts an influence on it. Suppose that *a*'s state is 0, *b*'s state is 1, $A_b = 2\ \mathrm{m}^2$, $P_{0\rightarrow 1} = 0.8$, $d_{ab} = 2.2$ m, and $\lambda_{0\rightarrow 1} = 0.6$. Suppose also that $\{\exists g | g(2,\ 0.8,\ 2.2) = 0.65\}$, then because $0.65 \geqslant 0.6$, a region of *a* will change its state to the state of *b*. The area of this region is calculated using the transition function of *a*. In figure 2(b) *a* is transformed into a raster format, and the number of cells corresponding to the area that must change state is calculated. Let us suppose that $\{\exists f | f(2,\ 0.65) = 0.21\ \mathrm{m}^2\}$ and the resolution of rasterization is 0.1 m, then $N_c = 0.21\ \mathrm{m}^2/0.1\ \mathrm{m}^2 = 21$. The cells of *a* that are nearest to the neighbor *b* are assigned the state *b* until the number of cells previously calculated is reached [figure 2(c)]. Then *a* is transformed back into a vector format and a new polygon is created with the area that has changed state being incorporated into *b* [figure 2(d)]. The topology of each polygon is updated after each geometric transformation.

2.2 Implementation

The VecGCA model was designed using the Oriented-Object Methodology and implemented in Java. Two additional libraries were used, OpenMap library (OpenMap, 2005) for the handling and display of shape files, and JTS Topology Suite (JTS, 2004) for the

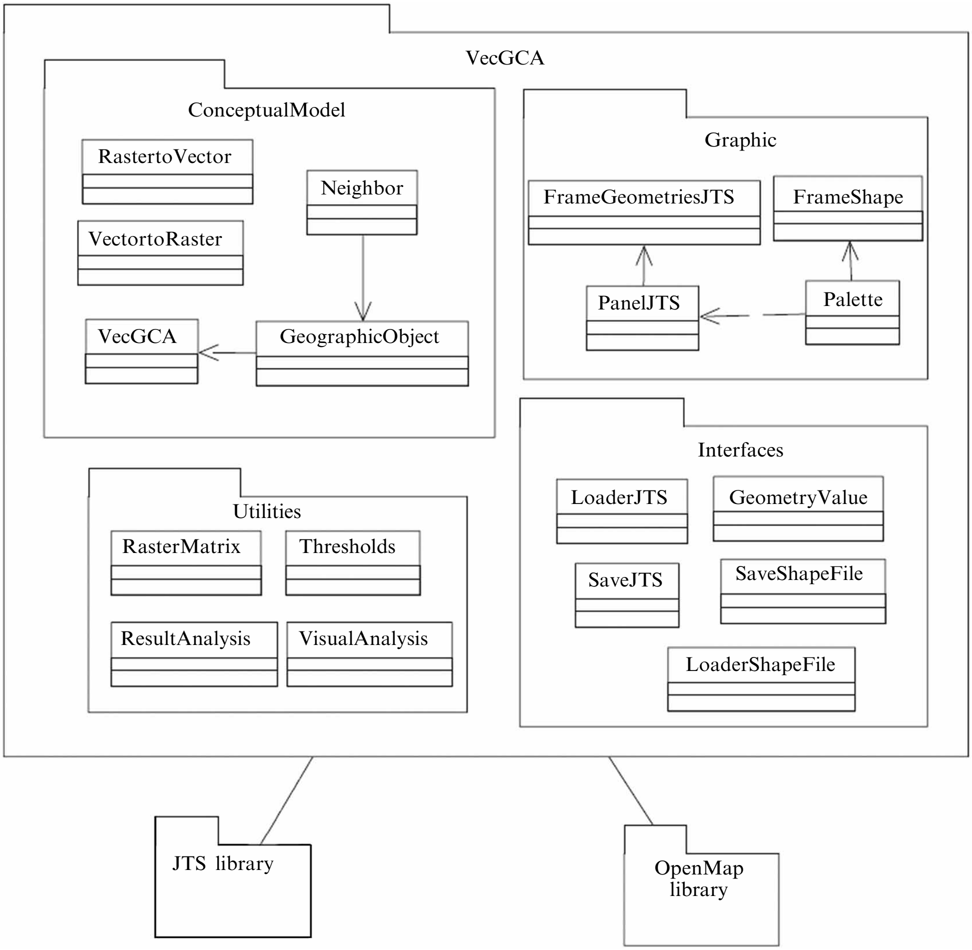


Figure 3. UML diagram of the library of components that define the VecGCA (vector-based geographic cellular automata) model.

handling of geometric objects (points, lines, polygons, polylines), buffer construction, and geometric operations (intersection, difference, union, and so forth).

The VecGCA model is presented as a library of software organized into four modules (see figure 3):

1. ConceptualModel. This includes the set of main classes composing VecGCA, the geographic space, and other additional classes that support the procedure of geometric transformations of polygons.
2. Graphic. This is a set of classes that displays the collection of geographic objects that define a space; these classes are implemented as subclasses of JFrame of the package java.swign.
3. Utilities. This includes additional classes that allow the handling of raster data and the calculation of the transition probabilities and thresholds from the comparison of two raster maps of different dates.
4. Interfaces. This contains classes that transform data from shape files to geographic objects and vice versa, using OpenMap and JTS libraries.

The two main classes that compose the model are VecGCA and GeographicObject. The VecGCA class allows the definition of the model that represents the system under study. It has a collection of attributes that document the model—its name, the description of the system, and the time unit that define the evolution of the system. Another set of attributes describes the characteristics associated with a space (defined as a collection of geographic objects), and the resolution used in the procedure of geometric transformations. The methods of this class allow the handling of all attributes, the main control of the simulation model, and the execution of the geometric-transformation procedure on geographic objects.

The GeographicObject class defines each geographic object of the VecGCA model, its behavior, its neighborhood, and the transition function. The attributes of this class store the information related to the actual state of the geographic object, its geometry, its neighborhood, and its neighbors; it also stores the information about its behavior associated with the possible changes, such as probabilities of transition and the threshold values. Some attributes are static (such as state, neighborhood, probabilities of transition and threshold), while others are updated through the simulation (neighbors and geometry). The methods of this class update the dynamic attributes and execute the evolution of the geographic object.

The procedure of geometric transformations is supported by two additional classes—VectorToRaster and RasterToVector. These classes execute the transformation of data from a standard vector format (shape file) to a standard matrix format (raster format), called rasterization, and from a matrix format to a vector format, called vectorization. Rasterization is a simple process that consists of overlaying a regular grid on the vector file and assigning to each cell the value associated with the polygon that contains it. The algorithm used in this model is the scan-line algorithm, commonly applied in computer graphics to convert vector maps to raster images (Healey et al, 1998). The vectorization process is more complicated. It consists of extracting, from a raster image, sequences of vectors that represent polygons, lines, or isolated points. The algorithm used in this model is a variation of the algorithm presented by Parker (1988) where the vectors are constructed in the border of each cell producing a stair effect in the polygons, but reducing the border and overlay problems between the adjacent polygons and preserving the integrity of the geographic space.

3 Methodology

The VecGCA model was tested using real data to simulate the land-use changes in an agroforested region in southern Quebec, Canada. The results were compared with those obtained with a raster-based CA model applied to the same study area under the same conditions.

3.1 Study area and dataset

The study area is the Maskoutains regional county municipality that covers an extent of 1312 km^2. The fertility of the land and its proximity to the Montreal city market create a situation that is highly favorable for agriculture but also generates high pressure on the remnants of forest in the region. Between 1999 and 2002 the region lost 23 km^2 of forested area, which corresponds to a decline of 10.5% (Soucy-Gonthier et al, 2003). The Maskoutains region was chosen for this study because of the availability of geographic data and the previous knowledge about the dynamics of land-use changes in this region (Ménard and Marceau, 2007), which facilitated the implementation and testing of the VecGCA model.

The main data sources are two land-use maps of the Maskoutains region originating from the classification of Landsat Thematic Mapper images acquired in 1999 and 2000 (Soucy-Gonthier et al, 2003). The images have an original spatial resolution of 30 m and the land-use classes are forest, agriculture, urban areas, and water (classes 1, 2, 3, and 4). These maps were transformed to a vector format using the RastertoPolygon conversion tool in ArcGIS 9.0 (ESRI, 2005).

3.2 Definition of the VecGCA model

To apply the VecGCA model to a specific study area, the user must define the neighborhood, the influence function, and the transition function that represent the dynamics of that study area. The neighborhood was defined as an external buffer of 30 m around each geographic object in order to establish a comparison with the Moore's neighborhood in the raster-based CA model with a cell size of 30 m that corresponds to the original resolution of the raster land-use maps.

The influence function is a function limited between 0 and 1, where 0 indicates no influence and 1 the greatest degree of influence. This function is directly proportional to the transition probability and the neighbor's area within the neighborhood, and inversely proportional to the distance between the centroids of the objects. For the model, this function is given by equation (4). The variables are the same as for equation (1).

$$g_{ab} = 1 - \exp[-P_{b\to a} A(t)_a / d_{ab}], \qquad 0 \leqslant g_{ab} \leqslant 1 \ . \tag{4}$$

Transition probabilities are defined from the comparison between the two land-use maps of 1999 and 2002. They are calculated according to equation (5).

$$P_{X\to Y} = \frac{A_{X\to Y}}{\sum_{i=1}^{4} A_{X\to i}} \ , \tag{5}$$

where $P_{X\to Y}$ is the transition probability from state X to state Y, $A_{X\to Y}$ is the total area that changes from state X to state Y.

Table 1 presents the transition calculated for a temporal resolution of three years.

The transition function that defines the area of change of each geographic object is given in equation (6).

$$f_b = \begin{cases} A(t)_a g_{ab} \ , & \text{if } g_{ab} \geqslant \lambda_{ab} \ , \\ 0 \ , & \text{otherwise} \ , \end{cases} \tag{6}$$

where λ_{ab} is the threshold value that represents the resistance of the geographic object b to change its state to the state of its neighbor a. This threshold value can be defined as the probability that a geographic object does not change its state to state X although all its neighbors are in state X. Equation (7) provides the threshold value:

$$\lambda_{ab} = 1 - P_{\max} \ , \tag{7}$$

where $P_{\max}$ is the probability that a geographic object b changes its state to the state of the geographic object a when all its neighbors are in the state of a.

Table 1. Transition probabilities defined from the comparison of the two vector land-use maps of the Maskoutains regional county municipality for 1999 (time t) and 2002 (time $t + \Delta t$) for a temporal resolution of three years.

t	$t + \Delta t$			
	water	forest	agriculture	urban
Water	1.00	0.00	0.00	0.00
Forest	0.00	0.78	0.22	0.00
Agriculture	0.00	0.02	0.98	0.00
Urban	0.00	0.00	0.00	1.00

Table 2. Threshold values indicating the minimum influence that a polygon must exert on another to produce a geometric transformation (where $t = 1999$; $t + \Delta t = 2002$).

t	$t + \Delta t$			
	water	forest	agriculture	urban
Water	1.00	1.00	1.00	1.00
Forest	1.00	1.00	0.11	1.00
Agriculture	1.00	0.37	1.00	1.00
Urban	1.00	1.00	1.00	1.00

Since the neighborhood configurations for irregular polygons are infinite, a simplified situation for the calculation of the threshold values was applied. All the polygons are considered to be squares of the same size, connected to each other as cells in the raster model. Then, $P_{\max}$ is the probability that a geographic object changes its state from state X to state Y when its eight neighbors are in state Y. This probability was calculated from the comparison of the two raster land-use maps (1999 and 2002) and it is given by the number of cells in state X in 1999 that had changed to state Y by 2002, while having eight cells in state Y in its neighborhood, divided by the total number of cells in state X with the same neighborhood in 1999.

Table 2 presents the threshold values calculated for the model. This procedure is implemented in the class Thresholds of the Utilities package and can be used to determine the threshold values in any model from two raster images of different dates. The threshold values for the transitions between the same states have been assigned the value 1, since they are not considered as changes of state.

The original resolution (30 m) of the available land-use maps was chosen for the rasterization procedure, in order to ensure that all objects can be represented in the raster format.

3.3 Definition of the raster-based CA model

Two stochastic raster-based CA models were also implemented to compare results with those obtained using the VecGCA model. These models differ only in their spatial resolution. First, a resolution of 30 m was chosen to provide a direct comparison with the VecGCA model, in which land-use data at their original resolution of 30 m are used to establish the initial conditions of the model. Second, a resolution of 100 m was used on the basis of earlier results obtained in a scale-sensitivity analysis conducted by Ménard and Marceau (2005), which indicates that 100 m is the best resolution for representing the dynamics of the study area.

Space was defined as a regular grid corresponding to a raster land-use map. Each cell can be in one of four states, namely forest, agriculture, urban, or water, but only two changes of state have been considered in this region, namely forest to agriculture and agriculture to forest. The Moore neighborhood was chosen to represent the influence of the adjacent cells on a cell. Probabilistic rules were calculated from the comparison between two land-use maps (1999 and 2002), according to the procedure described in Ménard and Marceau (2005), where a forested cell with n agricultural cells in its neighborhood has a probability of changing to agriculture equal to the number of forested cells with n agricultural neighbors that changed to agriculture between 1999 and 2002, divided by the total number of forested cells with the same neighborhood in 1999. The transition probabilities were calculated using raster land-use maps at 30 m (see table 3) and 100 m spatial resolution (see table 4).

Table 3. Transition probabilities for the raster-based cellular automata model of 30 m.

States		Number of neighbors whose state at time t is equal to the cell's state at time $t+\Delta t$								
t	$t+\Delta t$	0	1	2	3	4	5	6	7	8
F	F	0.096	0.175	0.195	0.280	0.424	0.547	0.602	0.683	0.880
F	A	0.000	0.326	0.411	0.475	0.590	0.741	0.818	0.841	0.929
F	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	F	0.000	0.111	0.174	0.265	0.357	0.498	0.528	0.507	0.638
A	A	0.799	0.770	0.678	0.677	0.713	0.795	0.864	0.910	0.991
A	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	U	0.882	0.888	0.920	0.918	0.944	0.986	0.990	0.996	1.000
U	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	W	0.900	0.954	0.979	0.910	0.888	0.938	0.988	0.998	1.000

Note. W = water; F = forest; A = agriculture; U = urban (where $t = 1999$; $t+\Delta t = 2002$).

Table 4. Transition probabilities for the raster-based cellular automata model of 100 m.

States		Number of neighbors whose state at time t is equal to the cell's state at time $t+\Delta t$								
t	$t+\Delta t$	0	1	2	3	4	5	6	7	8
F	F	0.196	0.342	0.478	0.541	0.650	0.762	0.804	0.868	0.928
F	A	0.000	0.142	0.208	0.260	0.368	0.481	0.555	0.693	0.832
F	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	F	0.000	0.039	0.078	0.121	0.214	0.323	0.360	0.403	0.661
A	A	0.577	0.712	0.743	0.735	0.825	0.909	0.939	0.969	0.995
A	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U	U	0.923	0.952	0.951	0.979	0.990	0.991	0.997	0.997	1.000
U	W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	U	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W	W	0.980	0.964	0.978	0.951	0.959	1.000	1.000	0.000	0.000

Note. W = water; F = forest; A = agriculture; U = urban (where $t = 1999$; $t+\Delta t = 2002$).

The transition probability from forest to agriculture with no agricultural neighbors and from agriculture to forest with no forested neighbors has been assigned the value 0, because they do not correspond to any significant ecological phenomena. Since a pseudorandom-number generator was used to implement the probabilistic rules, ten replicates of each simulation were performed using the raster-based CA model and the mean results were considered in the comparison with the VecGCA model.

3.4 Model simulations

Two simulation periods were considered in the study:

1. From 1999 to 2002, in order to compare the results of the VecGCA model and the raster-based CA models with the 2002 land-use map. The 1999 land-use maps (in raster and vector formats) were used to establish the initial conditions in the raster-based CA models and the VecGCA model. A temporal resolution of one year was chosen and the transition probabilities for this temporal resolution were calculated from the transition probabilities of three years using the exponential method presented by Yeh and Li (2006) where the transition probability P calculated for a time step t is substituted by P^n for a time step T where $T = n \times t$.
2. From 2002 to 2032, in order to observe the performance of the VecGCA when executed over a longer period of time (10 time units) and to compare its results with the results obtained with the raster-based CA models. The 2002 land-use map (in raster and vector formats) were used to determine the initial conditions of each model. A temporal resolution of three years was used for the simulations.

Validation of the first simulation period was done by comparing the simulation outcomes of the VecGCA model with the vector land-use map for 2002, the 30 m raster-based CA model outcomes with the 30 m raster land-use map for 2002, and the 100 m raster-based CA model outcomes with the 100 m raster land-use map for 2002. The results for the second simulation period have not been validated since our purpose was to evaluate the performance of the VecGCA model in comparison with the raster CA models over a period of time.

In order to compare the simulation outcomes produced by the VecGCA model and the raster-based CA models, a landscape analysis using Fragstats 3.3 (McGarigal et al, 1995) was performed on the results of the raster-based CA models. The number of patches for each land use was calculated using the raster maps generated by the raster-based CA models and compared with the number of polygons produced in the VecGCA model.

4 Results

The simulation results obtained for the period 1999 to 2002 are presented, followed by a description of the possible states of the system in the following 30 years, according to simulations conducted for 2002 to 2032.

4.1 Simulation results for the period 1999 to 2002

The results produced by the VecGCA model and the raster-based CA models were compared with the vector land-use map for 2002 and the 30 m and 100 m raster land-use maps for 2002, respectively, to assess the capacity of the models to capture effectively the dynamics of the study area (see table 5). The three models reveal a proportion of land use very similar to the corresponding reference map, but the landscape configuration is quite different. With the 30 m raster-based CA model, the total number of patches increases from 3470 in 1999 to 5116 in 2002. This increase is due to the fragmentation of large patches into small forested patches produced by the change of state of forest cells to agriculture cells when at least one agriculture cell is present in the neighborhood. These changes are possible due to the high transition probabilities calculated from the comparison between the two 30 m land-use maps (1999 and 2002) (table 3). These results are in accordance with a previous scale sensitivity investigation conducted by Ménard and Marceau (2005) indicating that the resolution of 30 m produces an unrealistic fragmentation of the territory.

With the other two models, the total number of patches (polygons) decreases with time, producing a less fragmented landscape (table 5). This can be explained by the

Table 5. Statistics describing the simulation outcomes of the VecGCA (vector-based geographic cellular automata) and the raster-based CA (cellular automata) models, and comparison with the raster land-use map for 2002.

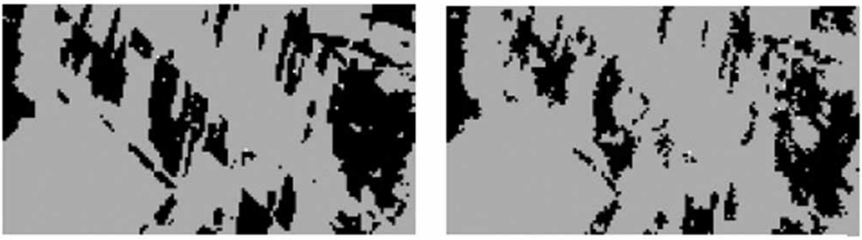
	Initial conditions land-use map 1999			Land-use map 2002			Simulation outcomes for 2002		
	vector format	raster format		vector format	raster format		VecGCA	raster-based CA	
		30 m	100 m		30 m	100 m		30 m	100 m
Proportion of forested area (%)	16.57	16.58	16.58	14.83	14.84	14.86	14.92	14.86	14.86
Proportion of agricultural area (%)	80.70	80.70	80.70	82.45	82.44	82.41	82.35	82.41	82.41
Number of patches (polygons) of forested area	1642	1453	915	1762	1335	707	1394	2727	871
Number of patches (polygons) of agricultural area	1865	902	218	2200	934	169	1650	1274	199
Total number of patches (polygons)	5247	3470	3118	5702	3387	1371	4666	5116	1569

Table 6. Results for the VecGCA (vector-based geographic cellular automata) model and the 100 m raster-based CA (cellular automata) model showing the proportion of simulated area that coincides with the state of the system in 2002 for each land use.

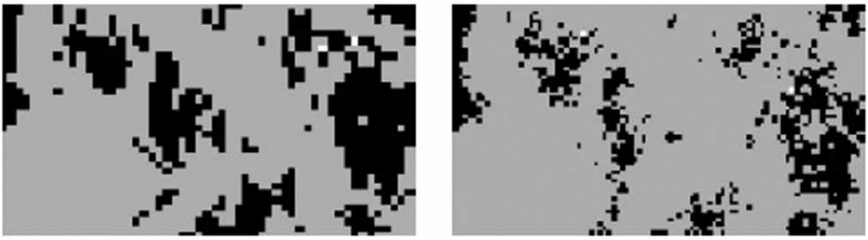
Land use	Proportion of simulated land use (%)	
	VecGCA	raster-based CA (cell size = 100 m)
Forest	100.00	77.47
Agriculture	99.00	93.04
Other	98.54	97.86
Average	99.18	89.46

disappearance of forested patches that are absorbed by large agricultural patches. In the 100 m raster-based CA model, this behavior is associated with the high transition probabilities for forest to agriculture when the number of agricultural neighbors is higher than five. In the VecGCA model, the disappearance of forested patches that are absorbed by large agricultural patches is explained by the high pressure that is put on a forested object with an agricultural neighbor that covers its entire neighborhood.

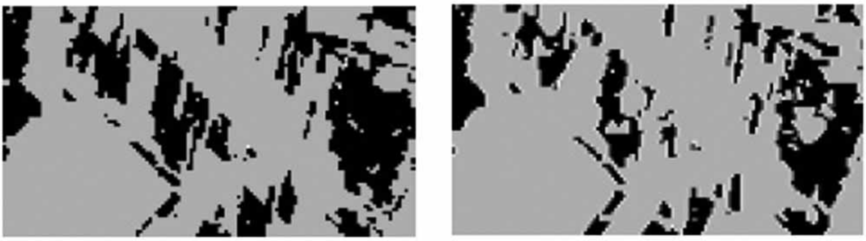
However, a difference between the results produced by the VecGCA model and the 100 m raster-based CA model can be observed when the 2002 vector and 100 m raster land-use maps with the corresponding simulation outcomes are overlaid (see table 6). An average of 99% of the landscape produced by the VecGCA model coincides with the land-use spatial distribution in the study area. In comparison, only 89% of the simulation outcomes of the 100 m raster-based CA model coincide with the land-use patches present in the reference map. All of the polygons of forest produced by the VecGCA model coincide with the real forested patches in the study area, in comparison to only 77% generated by the 100 m raster-based model. These results can be explained by the capacity of the VecGCA model to reproduce the evolution of the



30 m raster land-use map for 2002 30 m raster-based CA result for 2002



100 m raster land-use map for 2002 100 m raster-based CA result for 2002



Vector land-use map for 2002 VecGCA result for 2002

Figure 4. The land-use maps for 2002, and the land-use spatial distribution generated by the VecGCA (vector-based geographic cellular automata) model and the raster-based CA (cellular automata) model for 2002.

objects by their change of shape, whereas the patches produced by the raster-based CA model are created by the agglomeration of individual cells changing state. The outcomes of the VecGCA model, obtained from the original spatial distribution of the land-use map, present a spatial distribution that is more similar to the state of the landscape in 2002 than the outcomes produced by the raster model, although the cell size of 100 m was identified as the most appropriate during a previous sensitivity analysis (Ménard and Marceau, 2005).

A zoomed-in view of a small portion of the study area illustrates the detailed spatial distribution produced by the VecGCA model and the raster-based CA models (see figure 4).

The landscape generated by the VecGCA model is characterized by large patches of well-defined boundaries, in comparison with the diffuse boundaries and a high landscape fragmentation produced by both raster-based CA models.

4.2 Simulation results for the period 2002 to 2032

The simulation results for the period 2002–32 reveal that the three models generate a similar trend, namely a decrease in forested areas and an increase in agricultural areas (see figure 5), but in different proportions. With the VecGCA model, the decrease in forested areas is low in comparison to the decrease shown in both raster-based CA models. Table 7 shows that with the VecGCA model the proportion of forested areas is

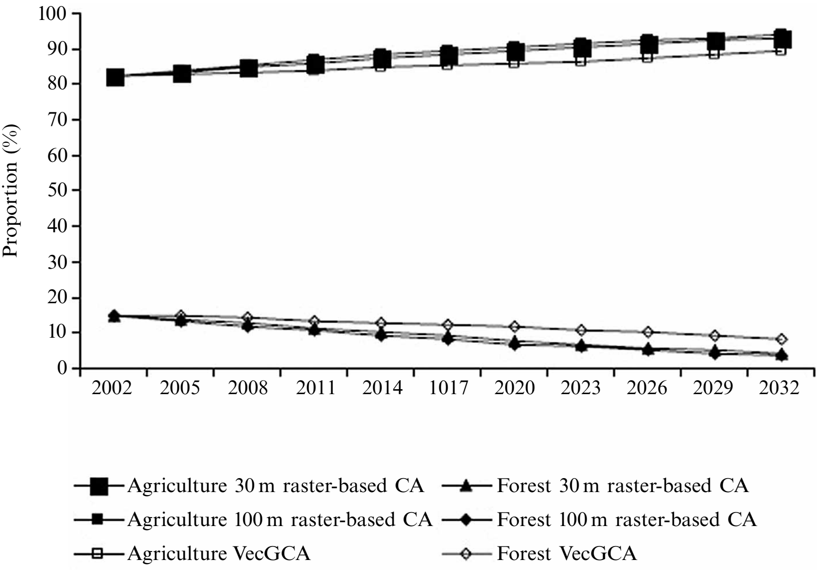


Figure 5. Proportion of forested and agricultural area calculated on the simulation outcomes of the VecGCA (vector-based geographic cellular automata) and the raster-based CA (cellular automata) model with resolutions of 30 m and 100 m for the period 2002 to 2032.

Table 7. Statistics describing the simulation outcomes of the VecGCA model and the raster-based CA models for 2032.

	Initial conditions land-use map 2002			Simulation outcomes for 2032		
	vector format	raster format		VecGCA	raster-based CA	
		30 m	100 m		30 m	100 m
Proportion of forested area (%)	14.83	14.84	14.86	7.58	4.30	3.58
Proportion of agricultural area (%)	82.45	82.44	82.41	89.69	92.98	93.69
Number of patches (or polygons) of forested area	1762	1335	707	1659	9753	474
Number of patches (or polygons) of agricultural area	2200	934	169	1359	628	83
Total number of patches (or polygons)	5702	3387	1371	4758	11499	1051

reduced from 14.83% in 2002 to 7.58% in 2032, whereas with the raster-based CA models the reduction is from 14.84% to 4.30% for the 30 m model and from 14.86% to 3.58% for the 100 m model. These results can be explained by the fact that with the VecGCA model, the decrease in forested areas is determined by the change of state in only a portion of the forested polygons produced by the influence exerted by the agricultural polygon neighbors at each time step. In comparison, in both raster-based CA models, the decrease in forested area is associated with the change of state of forested cells in the totality of their surface.

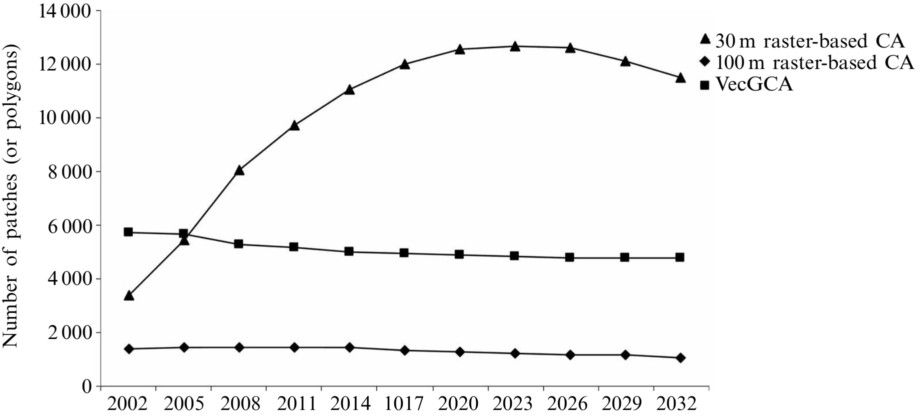


Figure 6. Total number of patches (or polygons) calculated on the simulation outcomes of the VecGCA (vector-based geographic cellular automata) and the raster-based CA (cellular automata) model with resolutions of 30 m and 100 m.

However, the landscape configuration produced by each model is different as revealed by the total number of patches or polygons (see figure 6). Table 7 shows that with the 30 m raster-based CA model, the total number of patches increases considerably from 3387 in 2002 to 11 499 in 2032, while it decreases slightly with the other two models. The 30 m raster-based CA model generates a considerably more fragmented landscape than the one produced by the 100 m raster-based CA model and the VecGCA model. The fragmentation in the 30 m raster-based CA model is due to the division of large forested patches into small ones, produced by the change of state of forest cells to agriculture cells when at least one agriculture cell is present in the neighborhood. For the other two models, the reduction in the number of patches (or polygons) is due to the disappearance of forested patches (or polygons) absorbed by large agricultural patches (or polygons).

The models display the same behavior for both simulation periods. The 30 m raster-based CA model produces a highly fragmented landscape that does not realistically represent the dynamics of the study area. These results are in agreement with those obtained in a previous study (Ménard and Marceau, 2005) indicating that 30 m is not an adequate cell size for simulating the dynamics of the Maskoutains region and that the better cell size is 100 m. The 100 m raster-based CA and the VecGCA models both capture well the dynamics of the study area. However, there is no need to perform a cell size sensitivity analysis with the VecGCA model since the polygons correspond to the real distribution of land-use patches in the landscape.

5 Discussion

The results obtained reveal that the VecGCA model produces an adequate evolution of the geographic objects, generating realistic spatial patterns when compared with the reference land-use map. These results are scale-independent and no sensitivity analysis is required to determine either the initial spatial distribution or the cell size of the model. However, some limitations are present in the model.

First, the rasterization used in the geometric transformation procedure introduces two problems. The selection of the cell size is arbitrary and affects the simulation outcomes. In addition, the rasterization produces a stair effect on the polygons' borders, generating errors in the delineation of the objects. A solution to overcome this problem is to replace the rasterization by a procedure that uses only vectors, where the portion removed from a geographic object is an irregular polygon whose borders conserve the initial definition of the object. Work is in progress to implement this solution in an improved version of the VecGCA model.

An issue that is not addressed in the VecGCA model is the possible sensitivity to neighborhood definition. Several questions are being investigated currently. Is the VecGCA model sensitive to neighborhood size? How to determine the appropriate neighborhood size for a specific study area? Should the neighborhood size around a geographic object be constant or variable?

Finally, the appearance of new isolated polygons simulating the emergence of land-use patches is not included in the VecGCA model. Nevertheless, in this version of the model new polygons appear when a geographic object is divided into several objects due to the influence of its neighbors, or because a nonadjacent neighbor produces a geometric transformation.

6 Conclusions

Raster-based CA models are sensitive to spatial scale. Previous assessments of the impact of spatial resolution on these model simulation results indicate that the arbitrary selection of cell size can generate outcomes that do not represent the real dynamics of the system under study. The VecGCA model is a novel approach that attempts to overcome the size sensitivity by allowing the representation of space as a collection of geographic objects where each object corresponds to a real entity of irregular shape and size (for example, a forest, a city, or a lake). Each geographic object has its own geometric representation (a polygon) and can define its proper behavior according to a transition function that determines its geometric transformations through time.

The VecGCA model encompasses an innovative procedure of a rasterization and vectorization that executes the geometric transformations of the geographic object. This procedure changes the state of a portion of that geographic object whose area is defined by the transition function and whose location is determined by the proximity to the neighbors that exert an important influence on it. This geometric transformation procedure constitutes the main difference between the VecGCA model and other vector-based models that have been developed.

Based on the new provided definitions of neighborhood, transition function, and geometric transformations, the VecGCA model reveals an adequate evolution of the geometry of the objects, producing realistic spatial patterns similar to the reference land-use map. Its initial spatial distribution corresponds to the real patches of the study area. In comparison, when using a raster-based CA model, a sensitivity analysis must be conducted to determine the cell size that best represents the dynamics of the study area and to generate the initial spatial distribution associated with this cell size.

This first implementation of the VecGCA model suffers some limitations introduced by the rasterization in the geometric transformation procedure. Continuing research includes a new implementation where only irregular polygons are handled. The algorithm will also be optimized to reduce the computation time. Finally, a sensitivity analysis on the neighborhood size will be done to determine its influence on the simulation outcomes. If the VecGCA model is sensitive to the neighborhood size, appropriate solutions will be implemented to ensure that the VecGCA model becomes a complete solution to the scale sensitivity in raster-based CA models, including cell size and neighborhood configuration.

Acknowledgements. The authors are very grateful to the three anonymous reviewers for their constructive comments and suggestions. This project was funded by scholarships awarded to Niandry Moreno by the Organization of American States, the University of Montréal, and the University of Calgary, and by a National Sciences and Engineering Research Council research grant awarded to Danielle J Marceau.

References

Benenson I, Torrens P M, 2004 *Geosimulation: Automata-based Modeling of Urban Phenomena* (John Wiley, Chichester, Wussex)

Chen Q, Mynett A E, 2003, "Effects of cell size and configuration in cellular automata based prey – predator modeling" *Simulation Modelling Practice and Theory* **11** 609 – 625

Clarke K, Bass J, Riggan P, 1994, "A cellular automata model of wildfire propagation and extinction" *Photogrammetric Engineering and Remote Sensing* **60** 1355 – 1367

Clarke K C, Hoppen S, Crane M P, 1997, "A self-modifying cellular automata model of historical urbanization in the San Francisco Bay area" *Environment and Planning B: Planning and Design* **24** 247 – 261

Couclelis H, 1985, "Cellular worlds: a framework for modeling micro – macro dynamics" *Environment and Planning A* **17** 585 – 596

de Almeida C M, Batty M, Monteiro A M V, Câmara G, Soares-Filho B S, Cerqueira G C, Pennachin C L, 2003, "Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation" *Computers, Environment and Urban systems* **27** 481 – 509

Dietzel C, Clarke K, 2006, "The effect of disaggregating land use categories in cellular automata during model calibration and forecasting" *Computers, Environment and Urban Systems* **30** 78 – 101

ESRI, 2005, "ArcGIS 9. Desktop GIS", http://www.esri.com/software/arcgis/about/desktop.html

Favier C, Chave J, Fabing A, Schwartz D, Dubois M A, 2004, "Modelling forest-savanna mosaic dynamics in man-influenced environments: effects of fire, climate and soil heterogeneity" *Ecological Modelling* **171** 85 – 102

Flache A, Hegselmann R, 2001, "Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics" *Journal of Artificial and Social Simulation* **4**(4), http://jasss.soc.surrey.ac.uk/4/4/6.html

Healey R, Dowers S, Gittings B, Mineter M, 1998 *Parallel Processing Algorithms for GIS* (Taylor and Francis, London)

Hu S, Li D, 2004, "Vector cellular automata based geographical entity", in *Proceedings of the 12th International Conference on Geoinformatics – Geospatial Information Research: Bridging the Pacific and Atlantic* Ed. S A Brant (Gävle University Press, Gävle, Sweden) pp 249 – 256, http://www.hig.se/~bjg/geoinformatics/files/p249.pdf

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** 217 – 241

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

JTS, 2004, "JTS Technology Suite", Vivid Solutions, Victoria, British Columbia, http://www.vividsolutions.com/jts/jtshome.htm

Kocabas V, Dragicevic S, 2006, "Assessing cellular automata model behavior 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: Planning and Design* **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** 131 – 152

Li X, Yeh A G-O, 2001, "Calibration of cellular automata by using neural networks for the simulation of complex urban systems" *Environment and Planning A* **33** 1445 – 1462

Li X, Yeh A G-O, 2002, "Neural-network-based cellular automata for simulating multiple land use changes using GIS" *International Journal of Geographical Information Science* **16** 323 – 343

Li X, Yeh A G-O, 2004, "Data mining of cellular automata's transitions rules" *International Journal of Geographical Information Science* **18** 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

McGarigal K, Marks B, Ene E, Holmes C, 1995, "Fragstats", Department of Natural Resources Conservation, University of Massachusetts, Amsherst, MA, http://www.umass.edu/landeco/research/fragstats/fragstats.html

Ménard A, Marceau D J, 2005, "Exploration of spatial scale sensitivity in geographic cellular automata" *Environment and Planning B: Planning and Design* **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 automata" *Landscape and Urban Planning* **79** 253 – 265

O'Sullivan D, 2001a, "Exploring spatial process dynamics using irregular cellular automaton models" *Geographical Analysis* **33** 1 – 18

O'Sullivan D, 2001b, "Graph-cellular automata: a generalized discrete urban and regional model" *Environment and Planning B: Planning and Design* **28** 687 – 705

OpenMap, 2005, "OpenMapTM. Open systems Mapping Technology", BBN Technologies, http://openmap.bbn.com

Parker D C, Manson S M, Janssen M A, Hoffmann M J, Deadman P, 2003, "Multi-agent systems for the simulation of land-use and land-cover change: a review" *Annals of the Association of American Geographers* **93** 314 – 337

Parker J R, 1988, "Extracting vectors from raster images" *Computers and Graphics* **12** 75 – 79

Rietkerk M, Dekker S C, de Ruiter P C, van de Koppel J, 2004, "Self-organized patchiness and catastrophic shifts in ecosystems" *Science* **305** 1926 – 1929 (Review)

Shi W, Pang M Y C, 2000, "Development of Voronoi-based cellular automata—an integrated dynamic model for geographical information systems" *International Journal of Geographical Information Science* **14** 455 – 474

Soucy-Gonthier N, Marceau D, Delage M, Cogliastro A, Domon G, Bouchard A, 2003 *Détection de l'évolution des superficies forestières en Montérégie entre juin 1999 et août 2002 à partir d'images satellitaires Landsat TM* [Survey of the evolution of forested areas in the Monteregie region, June 1999 – August 2002, from Landsat-TM remote sensing images], report presented to the Agence Forestière de la Montérégie; copy available from the authors

Stevens D, Dragicevic S, Rothley K, 2007, "iCity: A GIS-Ca modelling tool for urban planning and decision making" *Environmental Modelling and Software* **22** 761 – 773

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

Takeyama M, Couclelis H, 1997, "Map dynamics: integrating cellular automata and GIS through Geo-Algebra" *International Journal of Geographical Information Science* **11** 73 – 91

Thiéry J M, d'Herbes H M, Valentin C, 1995, "A model simulating the genesis of banded vegetation patterns in Niger" *Journal of Ecology* **83** 497 – 507

Verburg P H, Nijs T, Eck J, Visser H, Jong K, 2004, "A method to analyse neighborhood characteristics of land use patters" *Computers, Environment and Urban Systems* **28** 667 – 690

White R, Engelen G, 2000, "High-resolution integrated modelling of the spatial dynamics of urban and regional systems" *Computers, Environment and Urban Systems* **24** 383 – 400

White R, Engelen G, Uljee I, 2000, "Modelling land-use change with linked cellular automata and socioeconomic models: a tool for exploring the impact of climate change on the island of St. Lucia", in *Spatial Information for Land-use Management* Eds M J Hill, R J Aspinell (Gordon and Breach, Amsterdam) pp 189 – 204

Wolfram S, 1984, "Cellular automata as models of complexity" *Nature* **311** 419 – 424

Wu F, 2002, "Calibration of stochastic cellular automata: the application to rural – urban land conversions" *International Journal of Geographical Information Science* **16** 795 – 818

Yeh A G-O, Li X, 2006, "Errors and uncertainties in urban cellular automata" *Computers, Environment and Urban Systems* **30** 10 – 28