A vector-based cellular automata model to allow changes of polygon shape

Niandry Moreno1,2, Danielle J. Marceau1
1 Geocomputing Laboratory, Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary
2500 University Drive N.W. Calgary (AB), Canada, T2N 1N4
2 Centro de Simulación y Modelos (CESIMO), Faculty de Ingeniería, Universidad de Los Andes
La Hechicera, Mérida 5101, Venezuela.
morenos@ula.ve, marceau@geomatics.ucalgary.ca

Keywords: Vector-based cellular automata, irregular polygons, polygon shape change, land-use change.

Abstract
In the last few years, cellular automata (CA) have been increasingly used to simulate geographic phenomena due to their computational simplicity and their explicit representation of space. However, recent researches have demonstrated that the classical raster-based CA models are sensitive to spatial scale. In attempts to overcome this problem, this paper presents a novel vector-based CA model, called VecGCA that defines space as a collection of geographic entities of different shapes and sizes that correspond to real-world entities. The model was tested with real data to simulate land-use changes in an agroforested area in southern Quebec, Canada. Its performance was assessed through visual and quantitative analyses of the shape and distribution of the spatial patterns that were generated when compared to the patterns produced by a conventional raster-based CA. The results obtained show that both models generate a similar trend in land-use change, but the landscape is considerably less fragmented with the VecGCA model compared to the raster-based CA model.

1. INTRODUCTION
Cellular Automata (CA) are systems of discrete space and time that can reproduce complex global patterns and behaviors from simple local interactions of cells (Wolfram 1984). A CA is defined by five basic elements: space, set of states, neighborhood, transition rules, and time. In the classic definition of CA, space is defined as an infinite and regular tessellation of cells of discrete states (generally a matrix). The set of states is the set of possible values associated to the cells. The neighborhood corresponds to a set of adjacent cells, and transition rules are usually specified as a rule table that defines the next state of the cell for each possible neighborhood configuration. The transition rules are uniformly applied to all cells at fixed time intervals.

Due to their computational simplicity and the similitude with the raster-based GIS (Geographic Information System) model, CA have been increasingly used to simulate land-use and land-cover changes (Almeida et al. 2003; Li and Yeh 2002; Ménard and Marceau 2005; Parker et al. 2003; 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 (Dietzel and Clarke 2005; Lau and Kam 2005; Li and Yeh 2000; Liu and Phinn 2003; White 1998).

However, recent studies have demonstrated that such raster-based CA models are sensitive to scale (cell size) and to the neighborhood configuration (Chen and Mynett 2003; Jantz and Goetz 2005; Jenerette and Wu 2001; Ménard and Marceau 2005). In a recent investigation of spatial scale sensitivity in a land-use change CA model, Ménard and Marceau (2005) advocate the development of alternative CA models, such as vector- or object-based models as a solution to this problem.

Lately, researchers have begun to implement vector-based cellular automata models, where space is represented as an irregular tessellation based on a Voronoi diagram, while the neighborhood relationships are defined by the Voronoi boundaries (Shi and Pang 2000; Shiyan and Dener 2004). But these models still limit the direct correspondence of the polygons with the real geographic objects and the explicit definition of the neighborhood relationships.

This paper presents a novel vector-based cellular automata model, called VecGCA, which overcomes the limitations of classic CA and of these recent implementations of vector CA. The proposed model was tested with real data to simulate land-use changes in an agroforested area in southern Quebec, Canada. Its performance was assessed through visual and quantitative analyses of the shape and distribution of the spatial patterns that are generated when compared to the patterns produced by a conventional raster-based CA.

In the following section, a detailed description of the proposed VecGCA model is presented that includes the conceptual model and the implementation details. Next, a description of the VecGCA model and the raster-based CA model is provided. Finally, results generated from both models are compared.

2. THE VecGCA MODEL
2.1. Conceptual model
The VecGCA model is an extension of the classical CA; therefore it is defined by the same five elements: space, set of states, neighborhood, transition rules, and time. However,
three components are altered from their basic definition: the space, the neighborhood and the transition rules.

2.1.1. Space
Space is defined as a collection of geographic objects of irregular shape, georeferenced, and whose spatial representation can be associated to a geometric feature (point, line or polygon). Each geographic object has its proper behavior and evolves through time according to a transition function that depends on the influence of its neighbors. A geographic object represents a real entity of the system under study, for example: a forested patch, a city, a lake, etc.

2.1.2. Neighborhood
The neighborhood is defined as the region of influence (external buffer) of each geographic object, and the neighbors are all geographic objects located within the region of influence (Figure 1).

![Figure 1. Neighborhood as defined in VecGCA. For the object a, the neighborhood defined by r includes the objects b, c, d, e, f, g and h](image)

Each neighbor exerts an influence on a specific geographic object. This influence is defined as a function that depends on the area of that neighbor within the region of influence, the distance between the centroid of the neighbor and the geographic object, and the probability that the geographic object changes state for the state of this neighbor. The value of this function varies between 0 and 1 and is given by Equation 1.

\[
g_{ab} = g\left( A(t)_a, P_{X_b(t) \rightarrow X_a(t+1)}, d_{ab} \right) \tag{1}
\]

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 \(t\),
- \(P_{X_b(t) \rightarrow X_a(t+1)}\) is the probability of transition from \(X_b(t)\) to \(X_a(t+1)\),
- \(X_b(t)\) is the state of the neighbor \(a\) at time \(t\), and
- \(d_{ab}\) is the distance between the centroid of the neighbor \(a\) and the centroid of the object \(b\).

2.1.3. Transition function
The transition function defines the area (area unit) that changes state in a geographic object for the state of each of its neighbors. It depends on the area of the neighbors within the neighborhood and its influence on the specific geographic object (Equation 2).

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

where
- \(f_b\) is the transition function of object \(b\).

This function has 0 as lower limit, when the influence of the neighbor is smaller than a threshold value (\(\lambda\)), and the total area of the geographic object as upper limit, when the whole area of the geographic object changes state. The threshold value represents the resistance of a geographic object to change state for the state of its neighbor. 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 \(X\). This 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 geometrical transformations. This procedure reduces the area of the geographic object (calculated by the transition function) from the region nearest to the corresponding neighbor.

2.2. Implementation
The VecGCA model was designed using the Oriented-Object Methodology (OOM) standard and implemented in Java. VecGCA uses two additional libraries: OpenMap library (OpenMap 2005) for the handling and display of shape files, and JTS Topology Suit (JTS 2004) for the handling of geometric objects (points, lines, polygons, polylines) and their topology.

The VecGCA model was designed and codified as a library of software organized into four packages (Fig. 2):
1. package vecGCA.conceptualmodel: it groups the classes of the conceptual model, the geographic space and other classes that support the geometrical transformations procedure;
2. package vecGCA.interfaces: it contains the classes that transform data from shape files to geographic objects and vice versa, using OpenMap and JTS libraries;
3. package vecGCA.graphics: it groups the classes to handle the visualization of maps and collection of geographic objects. These classes are implemented as subclasses of JFrame of the package java.swing;
4. package vecGCA.utilities: this package handles raster data and additional classes that calculate the transition probabilities and threshold values from the comparison of two raster maps of different dates.

![Figure 2. Package diagram of VecGCA library](image)

The VecGCA class is the main class of the library, included in the package VecGCA.conceptual model. An instance of this class defines a particular system under study. It has a set of static attributes that document the model (name, description, time unit, space resolution) and a dynamic attribute "the space". The methods of this class allow the handling of all attributes and the main control of the simulation model. Two important methods of this class are updateNeighbors() and updateNeighborsInfluences() which update the list of neighbors for all geographic objects and the influence value and the transition function for all neighbors of each geographic object, respectively. The methods rasterization() and cellsChange() support the procedure of geometrical transformations. These methods are called from the method simulation(), which controls the execution of the evolution of the model. Figure 3 presents a flow chart of the simulation process in the VecGCA model.

![Figure 3. Simulation flow chart of VecGCA model](image)

In addition, each geographic object is defined as an instance of the GeographicObject class. The attributes of this class store the information related to the actual state of the geographic object, its geometry, its neighborhood and its neighbors, transition probabilities and 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 updateInfluences() and updateTransitionFunction() methods update the influence value and the transition function value for each neighbor of the geographic object. For each instance of this class, these methods must be rewritten specifying the influence function and the transition function corresponding to the study area. The method rulesTransition() controls the call of these two methods.
rasterization is a simple process that consists of overlaying a regular grid on top of the vector file and to assign to each cell the value associated to 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 in 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 (Parker 1988). The detailed algorithm for the geometrical transformation is presented in Figure 4.

1. Discretize the minimum bounding box around the external buffer (neighborhood) of the geographic object (g).
2. Return a matrix \( a \) where the value of each cell is the geographic object ID corresponding.
3. For each g's neighbor (neighbor)
   a. If neighbor.area_change > 0 then
      i. Change in the value of the cells nearest to the corresponding neighbor to its ID value. Start by the adjacent cells to the geographic object and next by the cells to a distance of 1, 2, ...n cells until the sum of the area of all cells changed is equal to neighbor.area_change or until there are no more possible cells to change.
   b. If there are cells that have changed then
      i. Vectorize (from g) a new polygon corresponding to all the cells whose value is equal to g.ID.
      ii. Update g.geometry with the new polygon obtained in (3.1).
   c. For each g's neighbor (neighbor)
      i. If neighbor.area_change > 0 then
         a. Vectorize (from g) a new polygon corresponding to all the cells whose value is equal to neighbor.ID
         b. Join (union) the new polygon (obtained from 3.3.1.1) with the neighbor.geometry
         c. Update neighbor.geometry with the new polygon obtained in (3.3.1.3).
   d. Update the topology of g and all its neighbors.

Figure 4. Geometrical transformation algorithm

3. LAND USE CHANGE MODEL

3.1. Study area

The study area is the Maskoutains region, an agricultural area covering 1312 km², located in Southern Quebec, Canada. Data used for the study include two land-use maps originating from Landsat Thematic Mapper images acquired in 1999 and 2002, respectively (Soucy-Gonthier et al. 2003). The original spatial resolution of these images is 30 m and the land-use classes are forest, agriculture and other (urban areas, roads, and water). These maps were transformed in a vector format using ArcGIS 9.0 (ESRI 2005).

3.2. VecGCA for the Maskoutains region

3.2.1. Space

Space is defined as a collection of patches of forest, agriculture, water and urban areas. Each patch corresponds to a polygon of the vector land-use map of the study area.

3.2.2. Neighborhood

The neighborhood is defined as an external buffer of 30 m around each patch. The influence function is a function 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, limited between 0 and 1, where 0 indicates no influence and 1 the greatest degree of influence (Equation 3).

\[
g_{ab} = 1 - e^{-p_{X_{ab}(t)} \cdot X_{ab}(t+\Delta t)} \cdot A(t)_{a} / d_{ab} \quad (3)
\]

Table 1 presents the transition probabilities calculated for a temporal resolution of three years from the comparison of the two vector land-use maps of 1999 and 2002.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( t+\Delta t )</th>
<th>Water</th>
<th>Forest</th>
<th>Agriculture</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Forest</td>
<td>0.00</td>
<td>0.78</td>
<td>0.22</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.00</td>
<td>0.02</td>
<td>0.98</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Urban</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

3.2.3. Transition function

The transition function that defines the area of change of each geographic object is defined as in Equation 4.

\[
f_{i} = f_{2} = \ldots = f_{n} = \begin{cases} A(t)_{a} \cdot g_{ab} & \text{if } g_{ab} \geq \lambda_{ab} \\ 0 & \text{otherwise} \end{cases} \quad (4)
\]

where \( \lambda_{ab} \) is a threshold value that represents the resistance of the geographic object \( b \) to change its state for the state of its neighbor \( a \). Table 2 presents the threshold values calculated for the Maskoutains model.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( t+\Delta t )</th>
<th>Water</th>
<th>Forest</th>
<th>Agriculture</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Forest</td>
<td>1.00</td>
<td>1.00</td>
<td>0.11</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Agriculture</td>
<td>1.00</td>
<td>0.37</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Urban</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

3.3. Raster-based CA for the Maskoutains region

Two stochastic raster-based CA models were implemented to compare the simulation results with those
obtained using the VecGCA model. These models differ only in their spatial resolution, namely 30 m and 100 m. 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. A resolution of 100 m was also used based on the results previously obtained in a scale sensitivity analysis conducted by Ménard and Marceau (2005), which indicates that this is the best resolution to capture the dynamics of the study area.

The models are defined using the following parameters:
2. Set of states: forest, agriculture, water and urban.
4. Transition rules: probabilistic rules, generated from the comparison of the two raster land-use maps (1999 and 2002). The transition probabilities were calculated according to the procedure described in Ménard and Marceau (2005).

3.4. Model simulations
Two simulation periods were considered in the study. The first period, from 1999 to 2002, was used to validate the results of the VecGCA model and the raster-based CA model (with a cell size of 30 m) using the 2002 raster land-use map. The 1999 raster land-use map and the 1999 vector land-use map were used to establish the initial conditions in the raster-based CA model and the VecGCA model, respectively. A temporal resolution of one year was chosen. 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).

A second period, from 2002 to 2032, was used to compare the results of the VecGCA model with the results obtained with the raster-based CA models employed as forecasting models. The 2002 raster land-use map at 30 m and at 100 m (produced from the rescaling of the original 30 m land-use map) were used to determine the initial conditions in the raster-based CA models. The initial condition for the VecGCA model was the 2002 vector land-use map. A temporal resolution of three year was used for the simulations.

4. RESULTS
4.1. Simulation from 1999 to 2002
The results of both models reveal a proportion of land-use similar to the proportion observed in the 2002 raster land-use map, namely 14.84% of forested area and 82.44% of agricultural area (Table 3).

However, a visual comparison between the raster land-use map, the raster-based CA and the VecGCA simulation results shows that the outcome of the VecGCA model presents a spatial distribution more similar to the state of the system in 2002 than the outcome produced by the raster-based CA model (Figure 5). With the VecGCA model, the transition probabilities, producing a landscape that is less fragmented, characterized by large patches of well-defined boundaries. In comparison, the landscape in the raster-based CA model is more fragmented, with diffuse patch boundaries, and a larger number of patches.

Table 3. Proportion of land-use and number of patches/polygons in the outcomes of the VecGCA model, the raster-based CA model and the 2002 raster land-use map.

<table>
<thead>
<tr>
<th></th>
<th>VecGCA</th>
<th>Raster-based CA (cell size 30 m)</th>
<th>Raster-based CA (cell size 100 m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of forested area (%)</td>
<td>14.92</td>
<td>14.86</td>
<td>14.84</td>
</tr>
<tr>
<td>Proportion of agricultural area (%)</td>
<td>82.35</td>
<td>82.41</td>
<td>82.44</td>
</tr>
<tr>
<td>Number of patches/polygons of forested area</td>
<td>1394</td>
<td>2727</td>
<td>1335</td>
</tr>
<tr>
<td>Number of patches / polygons of agricultural area</td>
<td>1650</td>
<td>1274</td>
<td>934</td>
</tr>
<tr>
<td>Total number of patches / polygons</td>
<td>4666</td>
<td>5116</td>
<td>3387</td>
</tr>
</tbody>
</table>

Figure 5. Spatial distribution of the 2002 raster land-use map and the distribution generated by the VecGCA model in 2002 and the raster-based CA model (30 m).
4.2. Simulation from 2002 to 2032

The results for this simulation period reveal the same trend in the three models, VecGCA and raster-based CA of 30 m and 100 m. A decrease in forested areas and an increase in agricultural areas can be observed (Fig. 6), but in different proportions. With the VecGCA model, the decrease in forested area is 8.17 km\(^2\) in 2032, which represents a decline of 44.18%. Whereas, with the raster-based CA models, the decrease in forested area in 2032 with a cell size of 30 m is 13.84 km\(^2\), corresponding to a decline of 71.03%; with a cell size of 100 m, the decrease is 14.75 km\(^2\), corresponding to a decline of 75.61%. These results can be explained by the fact that with the VecGCA model, the decrease in forested areas is determined by the reduction of the forested patches as defined by a mathematical function with the precision allowed by the spatial resolution used during the rasterization procedure. With both raster-based CA models, the decrease in forested area is associated to a random factor that determines if a cell changes state in the totality of its surface.

Similarly to the previous simulation period, the spatial distribution and the total number of patches/polygons for each land use are different. The difference in the initial distribution of the number of patches in the raster-based models and polygons in the VecGCA model is due to the vectorization process executed to transform the raster land-use map into a vector land-use map.

The spatial distribution for the raster-model CA of 30 m is more fragmented than the spatial distribution produced by the raster-based CA model of 100 m and the VecGCA model. The number of forested patches in the raster-based CA model of 30 m increases from 1335 in 2002 to 9753 in 2032, whereas in the VecGCA, the number of forested polygon decreases from 1762 in 2002 to 1659 in 2032. In the raster-based CA model of 100 m, the number of forested patches decreases from 707 in 2002 to 474 in 2032 (Fig. 7).

5. CONCLUSION

A previous scale sensitivity investigation conducted by Ménard and Marceau (2006) indicates that the resolution of 30 m is not appropriate to simulate the dynamics of the study area and that the best resolution is 100 m. The results obtained with the raster-based CA model of 100 m and the VecGCA models are similar, which indicates that the VecGCA model represents well the dynamics of the study area.

The advantage of the VecGCA model is that its initial spatial distribution is the original distribution of the system under study (the 2002 vector land-use map), which corresponds to the real patches existing in the Maskoutains region. In contrast, with the raster-based CA model a scale sensitivity analysis must be conducted to determine the best resolution that represents the dynamics of the system under study and the initial conditions of the model.
based CA models. These results are due to the VecGCA model capacity to change state in only a portion of the area of a geographic object, whereas in the raster-based CA models the changes of state are executed in the totality of the cell area.

VecGCA is a more complex model than the raster-based CA model. Its implementation is computationally intensive due to the reconstruction of the topology after each geometric transformation of the polygons. Work is currently undergone to optimize the algorithm in order to reduce the computation time.

6. ACKNOWLEDGMENTS

This project was funded by a scholarship awarded to N. Moreno by OAS (Organization of American States) and University of Calgary, by a research grant awarded to D.J. Marceau by NSERC (Natural Sciences and Engineering Research Council).

7. REFERENCES


**Niandry Moreno** is a doctoral student in the Department of Geomatics Engineering at the University of Calgary, Canada. She holds a Bachelor degree in System Engineering (1997) and a M.Sc. Degree in Computer Science (2001) from University of Los Andes, Venezuela. She is an assistant professor of the Simulation and Modeling Center at the University of Los Andes, Venezuela from 2000. Her current research is focused on vector-based cellular automata models and the calibration of cellular automata models to simulate urban growth.

**Danielle J. Marceau** is a professor of the Department of Geomatics Engineering at the University of Calgary, Canada. She completed a M.Sc. degree at the University of Sherbrooke (Québec) in 1988, and obtained a Ph.D. in 1992 at the University of Waterloo (Ontario) in Physical Geography and Remote Sensing. She then worked as a researcher at l'INRS-Eau (Institut national de la recherche scientifique) in Québec City from 1992 to 1993. She was a professor at the Department of Geography, University of Montreal, from 1993 to 2005 where she established the Geocomputing Laboratory in 1997. Her research program is entitled *Environmental Geocomputation*, an emerging discipline dedicated to the solution of complex environmental problems using advanced geocomputational approaches. It is based on the theoretical foundations of Geographic Information Science, Complexity Theory, and Geocomputation and involves the development and integration of innovative approaches in remote sensing, geographic information systems, individual-based modeling (cellular automata and multi-agent systems), and geovisualization for environmental resource management.