UCGE Reports
Number 20265

Department of Geomatics Engineering

Semi-automated calibration of a cellular automata model to simulate land-use changes in the Calgary region

(URL: http://www.geomatics.ucalgary.ca/research/publications/GradTheses.html)

by

Jean-Gabriel Hasbani

January 2008

UNIVESITY OF CALGARY

Semi-automated calibration of a cellular automata model to simulate land-use changes in the Calgary region

by

Jean-Gabriel Hasbani

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

JANUARY, 2008

Abstract

This thesis describes a semi-automated interactive method that has been implemented to calibrate a cellular automata model developed to simulate land-use changes in the Calgary region, Alberta. Historical land-use maps are read and factors responsible for driving the land-use changes, such as the distance to the road network, are identified. A frequency histogram is produced for each combination of land-use changes, neighborhood configuration, and driving factor. This information is analyzed to automatically create the transition rules that can be applied for the simulation. This flexible and interactive calibration procedure allows a user to display the influence of each driving factor on past land-use changes and to select how this factor will influence the CA model when forecasting future land-use development. Multiple processes driving a land-use change are identified and simulated, and the application of constraints enables the simulation of "what-if" scenarios. The model generates realistic results in terms of land-use patterns and new urban development.

Acknowledgment

To Isabelle, my partner in life, for her constant support and happiness and for listening to all kind of programming problems and tricks. Explaining my problems often led to a solution! Thank you for this passive tremendous help!

To Danielle Marceau for her trust, dedication to science, for having provided invaluable advices and for setting research limits allowing this master thesis to be completed in a reasonable amount of time.

To the government of Canada for allowing me to spend the last 6.5 years studying in this country.

To Danielle Marceau and the department of Geomatics Engineering for providing students with a generous income, allowing us to dedicate all our time to research.

To the Calgary Regional Partnership for their trust, advices and funding of this project. A special thank-you is addressed to Colleen Shepherd, project manager, for her constant support and appreciation of this project since its early stages.

To my parents for their support, for their life long encouragement to always learn more and for giving me a set of values, including the application of sciences for the goodness of the commons.

To Chengqian Zhang for classifying the satellite images into land-use maps. Your hard work is at the foundation of this research. Our never ending discussions on GIS and remote sensing were both interesting and entertaining!

To Cheng, Fang, Juan, Mike, Morshed, Niandry and Nishad for all the interesting technical discussions on modeling.

Last but not least, thank you to you, the reader of this thesis!

Table of contents

List of figures

List of tables

List of Symbols and abbreviations

CA……………..Cellular Automata
CRP……………Calgary Regional Partnership
IDL…………….Interactive Data Language
M.D……………Municipal District
Mb……………..Megabyte
RI………………Resemblance Index

$\bar{x}_i$ …………….Mean value for layer *i*
σ_i …………….Standard deviation for layer *i*
$\in$ ………………Belongs to
$\Re^+$ …………….All positive real numbers

Chapter 1: Introduction

The population in Alberta grows faster than anywhere else in Canada. For the year 2006-2007, the Canadian population grew by 1.01% or 329,000 inhabitants while the Albertan population grew by 3.03% or 103,000 inhabitants (Alberta Finance Statistics, 2007; Statistics Canada, 2007a). Accordingly, the Elbow River watershed, located in southwest Alberta, is a fast growing area in which the spatial competition for different land uses is high. The development of new towns are being projected in and around the watershed, like the Harmony-Springbank residential project between Calgary and Cochrane with 3,500 new homes (M.D. of Rocky View, 2007) or Seebe, at the door of the Rocky mountains, with 5,500 new inhabitants (Remington, 2006). Several sectors of activities are competing for land and environmental resources including forestry, agriculture, urban development, oil and gas, recreation and conservation. Consequently, planners and politicians have to make hard choices that may have long-term effects on local communities but also on a broader region, in terms of natural resources, infrastructure, tax income, environment and quality of life (CRP, 2007a). The issue is becoming critical and the Albertan government, who abrogated the Calgary Regional Planning Commission in 1995, is now aiming at establishing a provincial land-use framework for "*effective management of competing land use interests*" (Gov. of Alberta, 2007; Municipal Affairs and Housing, 2007).

In order to help decision makers, tools designed to foresee the possible outcomes of intended land-use plans tackling the population and economic growth are required. Such tools should be designed to allow planners and politicians to understand the regional competition for land-use and to test the impacts of their envisioned solutions (Beynon *et al.*, 2002; Pettit, 2005). The Calgary Regional Partnership (CRP), a regional inter-municipal association incorporated in 2003 to address regional challenges including land-use planning is seeking such a decision support tool (CRP, 2006; 2007a; 2007b).

1.1 Cellular automata models

Cellular Automata (CA) models are appropriate tools for addressing such land-use forecasting needs. They have been first proposed by Ulam and Von Neumann in 1951 to study self reproducing artificial structures. Wolfram (1984) redefined them as dynamic, spatially explicit models having five main basic characteristics: 1) they have a regular 'infinite' discrete lattice of cells in one or two dimensions, 2) they have an internal clock representing a discrete time and cells are updated at each predefined temporal step, 3) each cell has a finite set of possible states, 4) transition rules are applied uniformly through time and space, and 5) the outcome of the transition rules, which is the new state of the cells, depends on the value of the cells in a local neighborhood.

In the 1990's, scientists realized the potential of CA models for environmental applications and have adapted them to simulate real-world spatio-temporal phenomena. Circular and extended neighborhoods are commonly used to reduce directional bias and better capture the spatial influence of neighboring cell states on the central one (Batty and Xie, 1994; White et al., 1997; Ward et al., 1999; Liu and Phinn, 2001; Torrens and Benenson, 2005).

Many authors apply a linear or non-linear distance function to extended neighborhoods to better capture the spatially dependent attractiveness or repulsiveness of a state over another as depicted in Figure 1.1 (White and Engelen, 2000; Liu and Phinn, 2001). Moreover, additional spatial factors are considered in the transition rules such as the distance to a main road or to the city center (Benenson, 2007). Authors have also used different representations of space (Torrens and Benenson, 2005), such as Voronoi regions (Shi and Pang, 2000) or cadastral units (Batty and Xie, 1994; Stevens *et al.*, 2007). Vector CA models using physically meaningful entities rather than their raster approximation have also been recently developed (Moreno and Marceau, 2006).

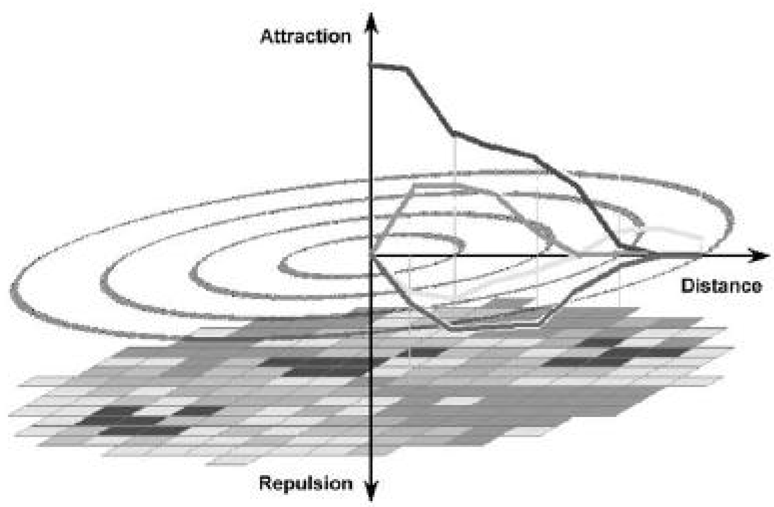

Figure 1.1. Distance influence in the neighborhood. The influence cells in the neighborhood have on the potential to change of the central cell. Reprinted from Straatman et al. (2004).

CA models are applied to a wide range of systems making them a great interdisciplinary tool (Parrott and Kok, 2000; White and Engelen, 2000). Among geographical applications, CA are used to simulate seismic activity (Georgoudas *et al.*, 2007), fire propagation (Ohgai *et al.*, 2007), prey-predator relationships (Chen and Mynett, 2003), urban development (Batty and Xie, 1994; Clarke et al., 1997; Torrens and O'Sullivan, 2000; Ward et al., 2000; Jantz et al., 2003; Cheng and Masser, 2004; Caruso et al., 2005; Fang et al., 2005; He et al., 2006), and land-use changes (Lett et al., 1999; Jenerette and Wu, 2001; Sui and Zeng, 2001; Li and Yeh, 2002; Soares-Filho et al., 2002; Bettingera et al., 2004; Dietzel and Clarke, 2007; Ménard and Marceau, 2007).

An example of a land-use CA model is depicted in Figure 1.2. Space is represented as a raster image where the cells can have one of four different land-use states: forest, water, agriculture or urban. For every cell, the model reads the state of the cells located within an extended neighborhood composed of thirteen cells and selects, according to the predefined transition rules, what would the new state of the cell be at the next time step, five years later. An example of a transition rule is:

IF land use is Agriculture

AND number of Urban cells in the neighborhood is equal or greater than 3

THEN change the land use to Urban

ELSE keep the land use as Agriculture.

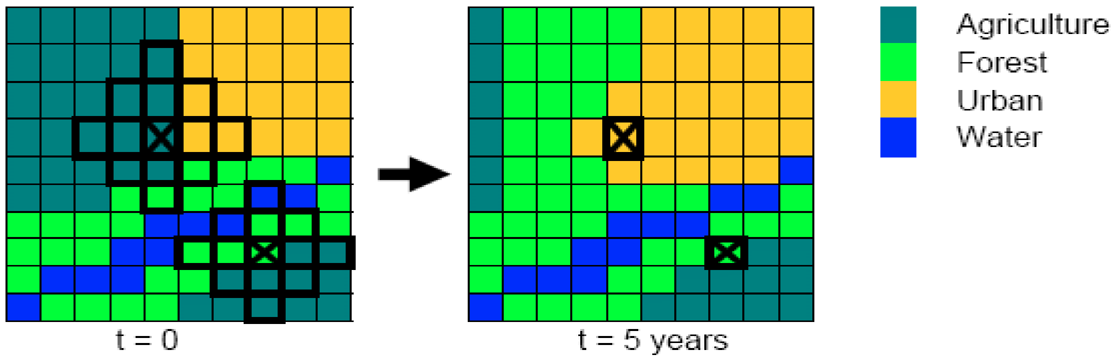

Figure 1.2. Example of a Cellular Automata. The cells in bold represent the neighborhood of the two cells marked with a cross.

1.1.1 CA models to study land-use and land-cover changes

CA models are particularly suitable for land-use change modeling (White and Engelen, 1993; Benenson and Torrens, 2004a) for several reasons. First, they are explicitly spatial. Secondly, the cell state is typically a land use but can be virtually anything, like population densities (Wu, 1998) or even a vector of values (White and Engelen, 2000). Moreover, such models can easily be constrained in various ways: it is possible to encourage, slow down or forbid land-use changes in a section of the territory in order to reflect local tendencies (Li and Yeh, 2000; Jenerette and Wu, 2001). 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 also be integrated into the CA in order to spatially allocate the land-use changes (White *et al.*, 1997). As a consequence, CA models are often designed to test “what-if” scenarios and policies (Ward *et al.*, 1999; White and Engelen, 2000; Yeh and Li, 2001; Wu, 2002; Jantz *et al.*, 2003; Li and Yeh, 2004) and are increasingly used by planners as a spatial decision support tool.

A key advantage of CA models is their simplicity. A CA is made of a limited set of equations and transition rules but can nonetheless exhibit complex behaviors (White and Engelen, 2000; O'Sullivan, 2001; Batty, 2005; Torrens and Benenson, 2005; Benenson, 2007). If well documented, the functioning of the CA model can be easily understood by non-expert users. Similarly, the model inputs and outputs are in most cases raster images ready to be used/created in any GIS/remote sensing software. The raster format also simplifies the definition of the neighborhood (White and Engelen, 2000; Benenson and Torrens, 2004b) and therefore simplifies the whole model. Consequently, it is easy to perform further analyses on the CA outputs, like quantifying the environmental impact of urban growth (Li and Yeh, 2000; Yeh and Li, 2001; Jantz *et al.*, 2003). CA can even be seen as an extension of a GIS in which dynamics are applied on the data (Takeyama and Couclelis, 1997; White and Engelen, 2000).

CA models are often described as powerful models to simulate complex systems (Benenson and Torrens, 2004b; Batty, 2005) as the output of the model emerges from the interactions between individual cells (Parwani, 2002). Generally, such outcomes can not be found by solving partial equations describing spatio-temporal interactions at local levels but only by running simulations (Wolfram, 1988; Parrott and Kok, 2000; Wu, 2000; Li and Yeh, 2004). CA models use simple rules to produce behavior of the greatest complexity (Wolfram, 1988; White and Engelen, 1993; Parrott and Kok, 2000; Shi and Pang, 2000; O'Sullivan, 2001; Benenson and Torrens, 2004a; Li and Yeh, 2004).

CA models are good for capturing patterns of land-use changes rather than precise location of changes (Jenerette and Wu, 2001; Jantz *et al.*, 2003; Jantz and Goetz, 2005). When decisions of changing a land use are made, people usually don't have all the available information at hand and therefore they make a choice that may not be the 'optimal' or expected one, introducing a degree of unpredictability in the system (White and Engelen, 1993; Parrott and Kok, 2000). For that reason, a stochastic factor is often included in the model: if the probability of a land-use change of a cell is greater than a random value, then the change will occur (White and Engelen, 2000).

1.1.2 Calibration of the model – defining the transition rules

Despite their simplicity, the challenge when implementing a CA model is the so called inverse problem, when the transitions rules have to be established according to temporal snapshots of the state of the model (Gutowitz, 1990) such as historical land-use maps. The calibration step involves finding the parameters of the predefined transition rules and the numerical values of these parameters so that the rules closely correspond to the land-use change processes reflected in the historical data. Consequently, the rules are based on an intuitive understanding of the processes as there is no obvious way of finding which parameter should or should not be included in the model (Wu, 2002). There are two types of transition rules referred in the literature: conditional and mathematical rules. As it will be done in this study, a rule can also be expressed as a combination of both types.

In the conditional form, the characteristics of the cell such as its distance to a road and the neighborhood configuration are taken into account in a series of *If… Then … Else* statements, as shown in the following example:

```
If        number_of_urbanized_neighbors > 4
          And distance_to_road < 1 km
Then      change_cell_state(urban)
Else if   number_of_commercial_neighbors > 5
Then      change_cell_state(commercial)
Else      //nothing, keep the same state
End
```

The mathematical form consists of equations that computes a probability of change from one land-use to another, where weights give more or less importance to each factor, as shown in the following example where P(x_to_y) is the probability that a cell changes from state x to state y:

P(agricultural_to_urban) = 0.95*(distance to road) + 0.05 * (number of urbanized neighbors) - 0.85 * (number of industrial neighbors)

P(urban_to_commercial) = - 2* (distance to road) + 1*(density of the neighborhood) - 0.8 * (number of existing business)

change_cell_state(**max**(P(agricultural_to_urban) ; P(urban_to_commercial)))

1.1.2.1 Conditional transition rules

Conditional transition rules are appealing for their simplicity. However, finding these rules may be complicated. There are only two methods referred in the literature. The first one relies on trial and error where the user modifies the rules slightly until a good result is produced (Wu, 2002). The second method, called data mining, is an automated method that produces a set of descriptive rules or a decision tree ready to be used as transition rules. The algorithm defines thresholds in the composition of the neighborhood and for the driving factors, which are additional values about each cell such as the land value, the distance to a main road etc., to maximize the likelihood that a given cell configuration leads to the correct type of land-use change (Li and Yeh, 2004). This method is promising as explicit rules are automatically derived. However, the authors did not consider the exact state of the cells within the neighborhoods, but rather the number of “developed” cells. Moreover, a given configuration can lead to one type of land-use change only, within a certain level of confidence, but does not acknowledge the fact that it may as well lead to several other types of land-use changes or to no change at all.

If properly established, conditional rules are good for modeling in the future provided that the evolution of the landscape is fairly similar to what was present in the historical data, that is, that the relationship between spatial variables and land-use changes do not vary over time (Li and Yeh, 2004). However, the model is not capable of adapting itself to new situations. As an example, in the case of a territory in which past settlements were very scattered, the model would not be able to simulate development of communities of people living nearby each others, even if it is the only way of spatially allocating the newcomers.

1.1.2.2 Mathematical transition rules

Mathematical rules are easier to obtain but are difficult to interpret. This method requires finding weights for each parameter of the transition rules. Without a proper calibration, the model is worth close to nothing as it will not be reliable (Wu, 2002). The rules and their underlying processes used to determine where a change will occur have to be accurate, as calibrating the model with some false assumptions will lead to reproduction of these false facts (Wu, 2002).

For example, the same rule can be applied to every land-use change:

Probability of change = factor1 + factor2 + … + factorN

However, after calibrating the model, the rules would be:

P(agricultural_to_urban) = a1 * factor1 + a2 * factor2 + … + aN*factorN

P(urban_to_commercial) = b1 * factor1 + b2 * factor2 + … + bN*factorN

Where the coefficients a1 to bN belong to $\Re$. In most CA models, the influence of the factors is additive (Straatman *et al.*, 2004), except in one case that we will see later in this section.

There are several ways to calibrate a model using mathematical transition rules. The first way is, again, by trial and error (Li and Yeh, 2004). The user will change some coefficients, also called weights, and will visually or quantitatively compare the simulation output to a real land-use map. This method is obviously time consuming and can be very cumbersome as soon as many variables are considered. Wolfram (1983) found that for N cell states and a neighborhood composed of K cells, there are up to N^{N^K} possible transitions, which reinforce the need for automatic calibration.

A second approach is to train a neural network to classify the land-use changes (Li and Yeh, 2002). A neural network is composed of layers, each of them having virtual neurons. The first layer of neurons corresponds to the inputs, which are in Li and Yeh's study the number of cells of each land-use type in the neighborhood of a cell and the value of the driving factors. There is also an output layer with a neuron for each possible new land-use state. Between these two layers, one or more hidden layers are used. Connections are made between the neurons of the different layers and each neuron computes a value that is sent to the next neuron. By using historical data, one can train the neural network so the correct connections are established. When simulating land-use changes, the neural network will apply these connections in order to find the most probable new land use. Li and Yeh (2002) obtained good results with this method but it function as a black-box where the user has neither control nor information on the mathematical equations used to determine the transition rules nor on the transition rules themselves. In addition, the rules might have no real geographical meaning (Wu, 2002; Straatman *et al.*, 2004).

The third way to calibrate a CA model with mathematical rules is to use a logistic regression. Wu (2002) successfully used this approach. In his model, each cell in the historical data is tagged as a binary value corresponding to “change to urban” or “no change to urban”. Then, statistical software compiles the values of the driving factors and the number of cells of each state in the neighborhood and computes the logistic regression. The resulting equation can then be applied during the simulation to compute the probability of “change to urban” according to the driving factors and the neighborhood composition. This approach has some limitations. An average of the historical neighborhood configurations leading to a change is considered and therefore all the variability is lost (Verburg *et al.*, 2004). Also, there is no distance function, which means that the attractiveness or repulsiveness of surrounding cells will remain constant in the neighborhood, no matter how far away it is from the central cell. Fang *et al.* (2005) also used a logistic regression but they considered the cross product of the factor values in order to reflect the interaction between the components of the system, which lead to a global accuracy of the CA of 80% versus 70% when not considering the cross product.

Brute force calibration can also be used. The model SLEUTH simultaneously uses five types of growth rules: spontaneous (new random development), spreading center (new development around sites developed by the spontaneous growth), edge or organic growth, and road influence (Messina *et al.*, 1999; Jantz *et al.*, 2003; Yang and Lo, 2003; Jantz and Goetz, 2005). This model relies on a brute force calibration, where the user modifies the value of each factor and compares the output with a known land-use map (Jantz *et al.*, 2003; Jantz and Goetz, 2005). To reduce the computational load, the calibration is done at coarse, medium and fine scales. The best set of parameters at a coarse level is refined at the medium and then at the fine scale level. The major problem with this technique is that a parameter could have a multimodal distribution in terms of output quality, and if a mode falls between two coarse values, the calibration algorithm will not see it. Also, this procedure still requires a human intervention to select the weights. The output validity is assessed by the calculation of thirteen indices, but the set of parameters giving the best result for one index could be the worst set when considering another index (Straatman *et al.*, 2004). Dietzel and Clarke (2007) found a way to combine these indices into a single metric, though they still have to compute millions of parameter combinations in order to find the best set. There is no information about the volume of data required to find the weights. Knowing this may greatly reduce the time of the calibration and possibly dismiss the coarse calibration step.

Straatman *et al.* (2004) proposed an automatic calibration of their model, which makes the CA model possibly reproducible by others. The first step is to estimate the relative values of the weights in order to avoid computing irrelevant parameter sets. Then, using the brute force method, they try to reduce the total error, expressed by the number of misclassified pixels. Once it is 'impossible' to reduce the total error, they adjust the parameter set to reduce the maximal error, which is within all neighborhoods the worst difference between the computed probability of change to the correct class and the highest probability of change in that cell (for another class, as if the cell is correctly classified, the local error is 0). This process is recursive, which means that once the maximal error is reduced, they compute again the total error and try to reduce it. The system stops when a predefined threshold is reached for either the total error or the maximal error. This method provides good results and is interesting as it can be universally applied. However, there is a major limitation: there is no unique solution and the ‘best’ set of parameters found is dependent on the random seed values, as the system always tries to reduce the error from the current set of parameters and always keeps the parameter sets that reduce the most the error and is therefore attracted by local minimum (Straatman *et al.*, 2004). The reproducibility of the calibration is therefore not guaranteed. The problem of the comparison index used is also present.

Verburg *et al.* (2004) describe an approach aimed at characterizing each land-use change. Since during the calibration different parameter sets can lead to the same pattern of change, the calibration is inappropriate for understanding and testing hypotheses concerning the underlying factors of urban development (Verburg *et al.*, 2004). Therefore, they look at the over or under representation of each land-use type in the neighborhood of the cells of each type of land use and find that different land uses have clearly distinct neighborhood characteristics. The main advantage of this method is that regional variation in the neighborhood characteristics are taken into account. However, this method does not generate transition rules (Verburg *et al.*, 2004).

An additional challenge when calibrating a CA model is related to the adequate selection of scale. The transition rule, the relation between a parameter and a process, is established at one spatial and temporal scale and may or may not be linearly translated at another scale (Messina et al., 1999; Jenerette and Wu, 2001; Candau, 2002; Kocabas and Dragicevic, 2004; Jantz and Goetz, 2005; Ménard and Marceau, 2005; Benenson, 2007), or can even have no equivalence at another scale (Railsback, 2001; Samat, 2006). Indeed, multiple analyses have been done to assess that CA models are sensitive to scale (Chen and Mynett, 2003; Ménard and Marceau, 2005; Benenson, 2007). To overcome this issue, one must either build a model that incorporates multiple scales (Jantz and Goetz, 2005), find a good relation between both spatial and temporal scales, or modify the transition rules accordingly to scale (Benenson, 2007).

All the previously described calibration methods suffer severe limitations. The transition rules are either difficult to obtain or to interpret. Some methods are a black box, others are too rigid to handle future important changes in the landscape as the weights in the transition rules remain constant through time (White and Engelen, 2000; Wu, 2000; Cheng and Masser, 2004). Most methods consider an average neighborhood configuration instead of multiple neighborhood configurations leading to a change and also assume that a given neighborhood configuration can lead to only one type of land-use change. Moreover, since the transition rules are hard coded into the model, it is often impossible to consider more land-use classes or features than the ones selected by the developer of the model. At last, some automated methods can be attracted by local minimum or a multi-scale calibration can result in hiding the best parameter set to the user.

1.2 Objective of the thesis

The objective of this research is to develop a new method to identify transition rules and calibrate a land-use CA model. Practically, the aim is to develop an accurate, predictable and repeatable method that can dynamically create transition rules based on the content of the historical data and accordingly compute parameters that define the processes driving the land-use changes. The proposed method will combine conditional and mathematical rules in order to have rules easy to understand by the user while benefiting from the adaptability of mathematical rules to new landscape configurations. This model should be very adaptable in terms of study area location, shape, number of land-use classes as well as in terms of spatial and temporal scale of the inputs. The model will allow the simulation of most types of development scenarios by combining different constraints. Most importantly, the calibration method will be easily understandable by a non-expert user and will also allow the user to understand the driving factors affecting land-use change within the study area by interacting with the model during the calibration.

This method will then be applied to simulate land-use changes in the Elbow River watershed from the past to the present to assess the accuracy of the model, and then from the present to the future in order to test “what-if?” scenarios of regional planning.

1.3 Organization of the thesis

The remainder of the thesis is organized as follow. The methods used to identify the transition rules, to calibrate the model and to apply the rules in a CA model are presented in

Chapter 2. Simulation results incorporating simple to more complex development scenarios are described in Chapter 3. The conclusions and recommendations for future work can be found in Chapter 4.

Chapter 2: Methodology

This chapter describes the study area and the main datasets used in the study, the architecture of the CA model and its main components. Then, the proposed calibration and simulation methods are explained and the constraints that are applied to simulate different scenarios are presented. The method used to validate the model is described in the last section.

2.1 Study area and datasets

The CA model has been designed to be applied to any study area and was tested on the Elbow River Watershed in southern Alberta (Figure 2.1). The Elbow River, which drains to the Bow River, is 120 km long and the watershed covers 1 200 km^2, with 5% covering the City of Calgary, 10% covering the Tsuu Tina nation, 20% covering the municipal district of Rocky View, and the remaining 65% covering Kananaskis country. The western zone of high mountains was not considered in this study as there are very limited land-use changes in this protected area. The elevation in the study area varies between 2 340 m and 1 010 m above see level (Figure 2.2 and Figure 2.3).

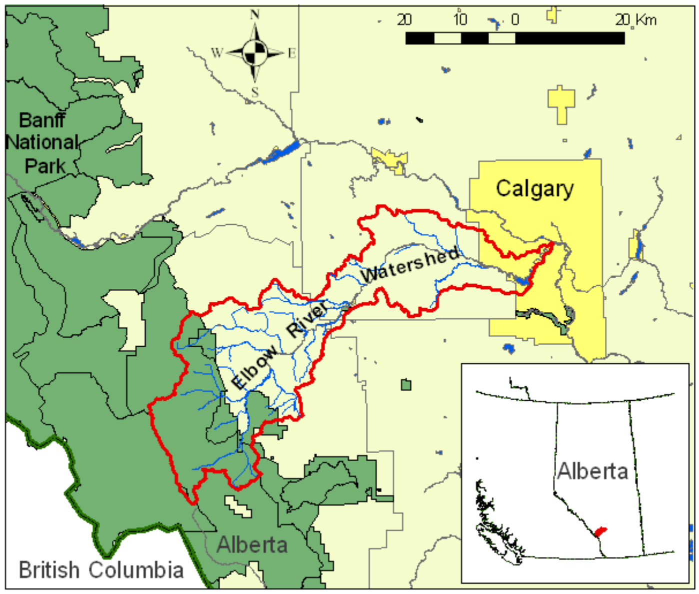

Figure 2.1. Location of the Elbow River Watershed.

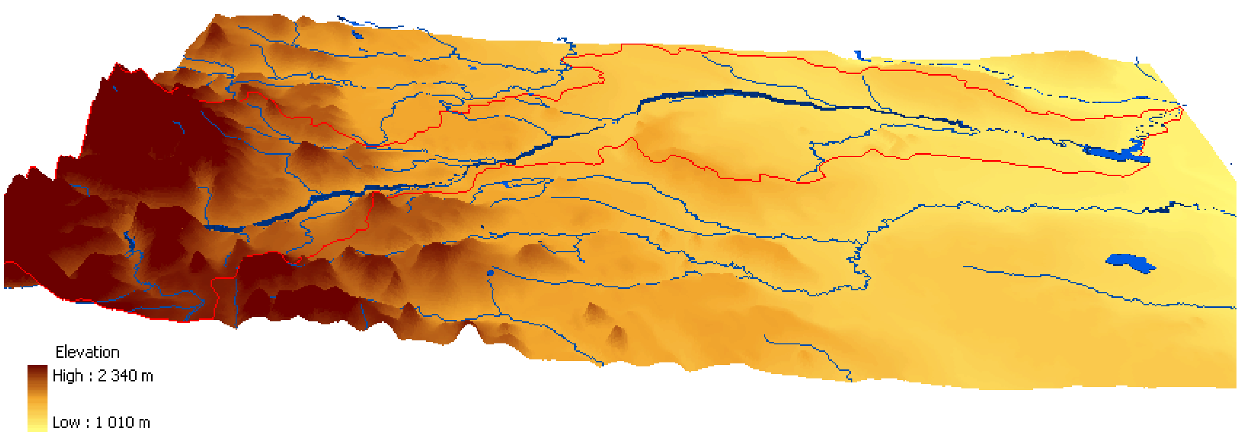

Figure 2.2. 3D representation of the Elbow river watershed.

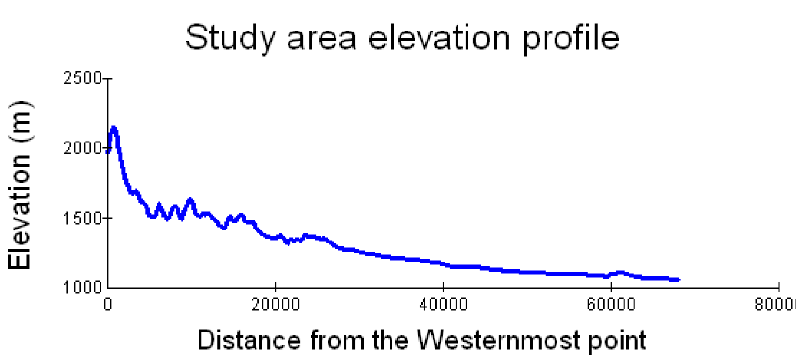

Figure 2.3. Elevation profile in the study area.

The watershed is characterized by large areas of forest (45.46%), agriculture and grassland (24.92%) and urban zones (6.32%). Calgary, the main city in the study area, is a fast growing city of one million inhabitants. The town of Bragg Creek, on the eastern portion of the watershed, is a typical rural town of 700 souls. Small farms and rural housing are also scattered within the watershed. In 2006, there were 94 000 inhabitants living within the study area (Statistics Canada, 2007b).

The historical land-use maps required for the CA model calibration were created by a remote sensing specialist using Landsat Thematic Mapper imagery acquired during the summers of 1985, 1992, 1996, 2001 and 2006. These maps are at 30 meter resolution and cover the whole watershed. This sequence of data allows the detection of trends in land-use changes over a period of 21 years and the capture of a variety of spatial processes influencing land-use changes. The selected years mainly correspond to years at which Canadian census data have been collected, permitting access to detailed population counts, though this information has not been used in the CA model.

The original thirteen land use classes were, for each date:

1. Water: water bodies including rivers, creeks, lakes and ponds
2. Road: principal and secondary roads
3. Rock: bare rocks located in the Rockies, as well as on low elevation ground
4. Forest: including conifer and deciduous stands, woods and shrubs
5. Agriculture: crop-on and harvested agricultural lands
6. Grassland: mostly located above the tree line, sometimes mixed with small shrubs
7. Parkland: vegetated lands mixed with trees, shrubs, and weeds

8. Construction: construction sites
9. Golf-Park: golf courses and parks
10. Clear-cut: Forested zones where most trees have been cut and removed
11. Urban areas
12. Cloud-Shadow: mixed with cloud-shadows and cliff-shadows
13. Tsuu Tina nation land.

Field verification has been done for the year 2006 and all maps have been shown to experts (scholar, CRP representatives and planners) for correction and validation.

These classes have been aggregated into five classes to reduce the computation time during the calibration and simulation. Table 2.1 provides the aggregation scheme. Pixels classified as Rock, Road or Cloud had been set to the value of the majority within a Moore neighborhood.

Table 2.1: Land-use aggregation scheme

Initial classes	**Aggregated classes**
Water	Water
Forest Clear-cut Golf-Park	Forest
Agriculture Grassland Parkland	Agriculture (Non-forested vegetation)
Urban areas Construction	Urban
Tsuu Tina land	Tsuu Tina land

A detailed analysis of the historical land-use maps exhibited some spatio-temporal inconsistencies due to classification and georeference errors. A computer program was written and applied to identify and correct these inconsistencies according to the following rules:

- Using an hydrology vector file created by Ensight Info, a company in Alberta specializing in producing base maps for the oil and gas industry, cells under a permanent river or lake have been set to Water.
- Using the same vector file, urban cells located within 50 m of a river in Calgary or within 70 m outside of Calgary have been changed to the state of the majority of the cells within the Moore neighborhood of these cells. This had to be done because for some years, the rocks on the shore of the river were misclassified as urban.
- Patches of water of less than three contiguous pixels were removed and the land use set to the majority within the Moore neighborhood.
- Within Calgary, if a cell was classified as urban at a given year, then it must remain urban for all the remaining years.
- Outside of Calgary, urban cells were changed to the state of the majority within the Moore neighborhood if they were not classified as urban at least half of the time in the remaining land-use maps. If they were kept as urban, then these cells would be classified as urban for all the remaining years in all the historical maps. For example, if a cell was Forest in 1985, Urban in 1992, Agriculture in 1996 and Forest in 2001 and 2006, then the land use of 1992 was changed to a new state, which was the state of the majority of the pixels within the Moore neighborhood. On the other hand, if a cell was Forest in 1985, Urban in 1992, Agriculture in 1996 and Urban in 2001 and 2006, then the land use for 1996 was set to Urban.

To reduce the computational load, the land-use maps were resampled at 60 m resolution using the nearest neighbor method (ESRI, 2005). Little information was lost during the resampling. The final land-use maps can be seen in Figures 2.4 to 2.8. The temporal distribution of the land uses is shown on table 2.2.

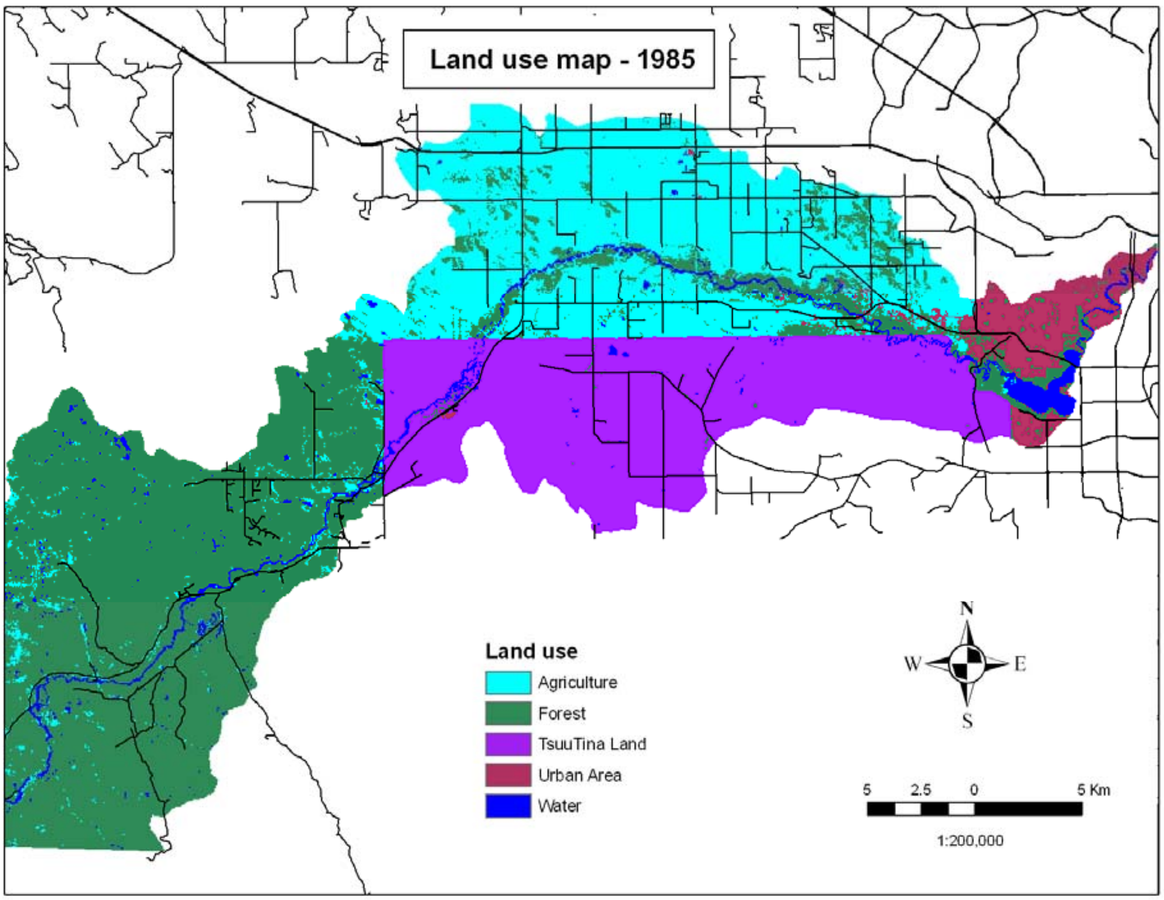

Figure 2.4. Land use map of the Elbow River Watershed - 1985

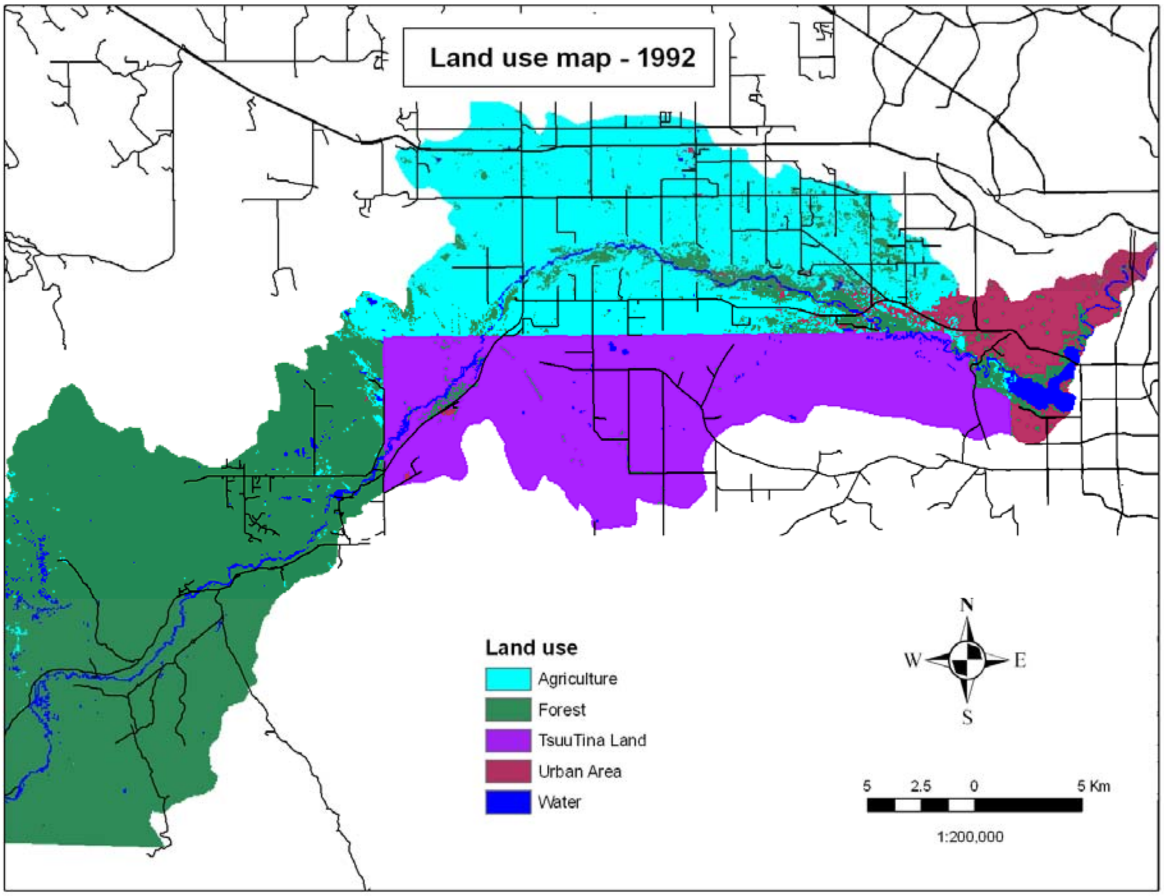

Figure 2.5. Land use map of the Elbow River Watershed - 1992

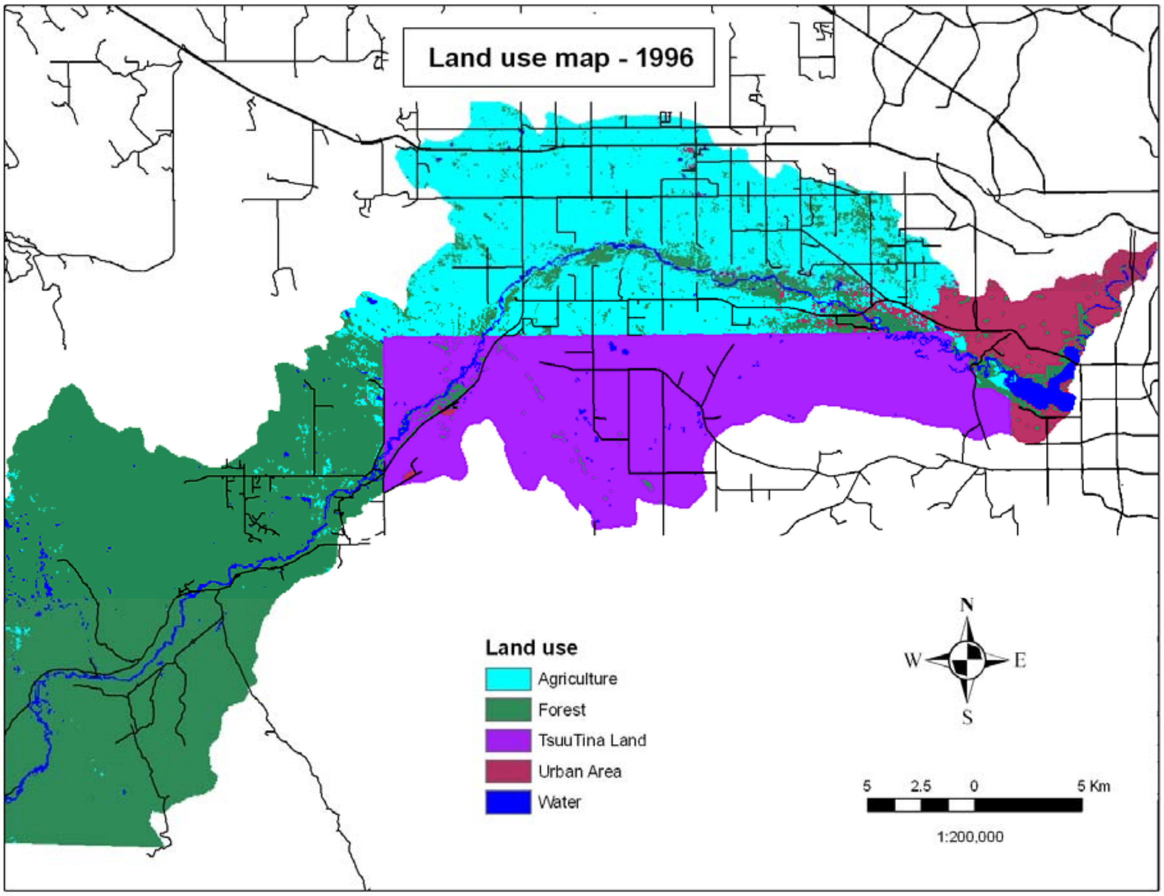

Figure 2.6. Land use map of the Elbow River Watershed - 1996

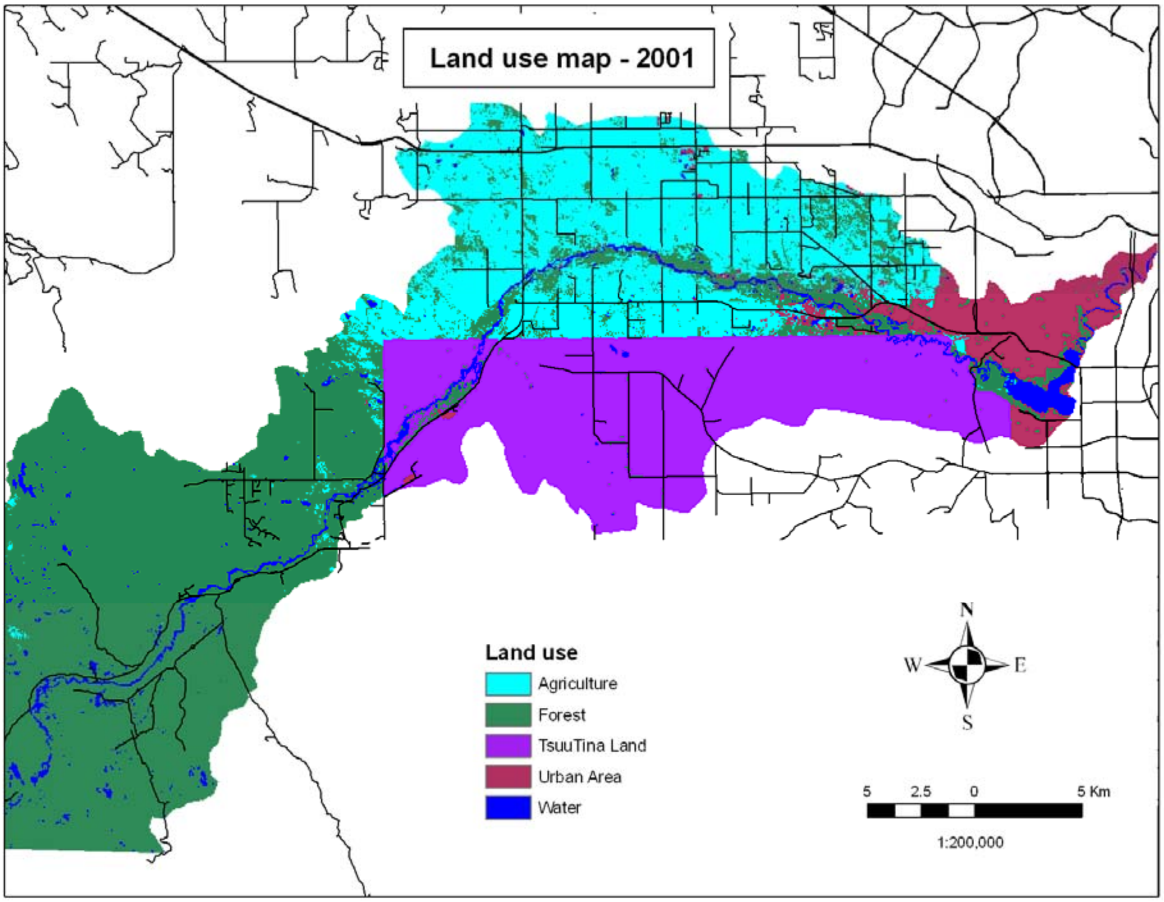

Figure 2.7. Land use map of the Elbow River Watershed - 2001

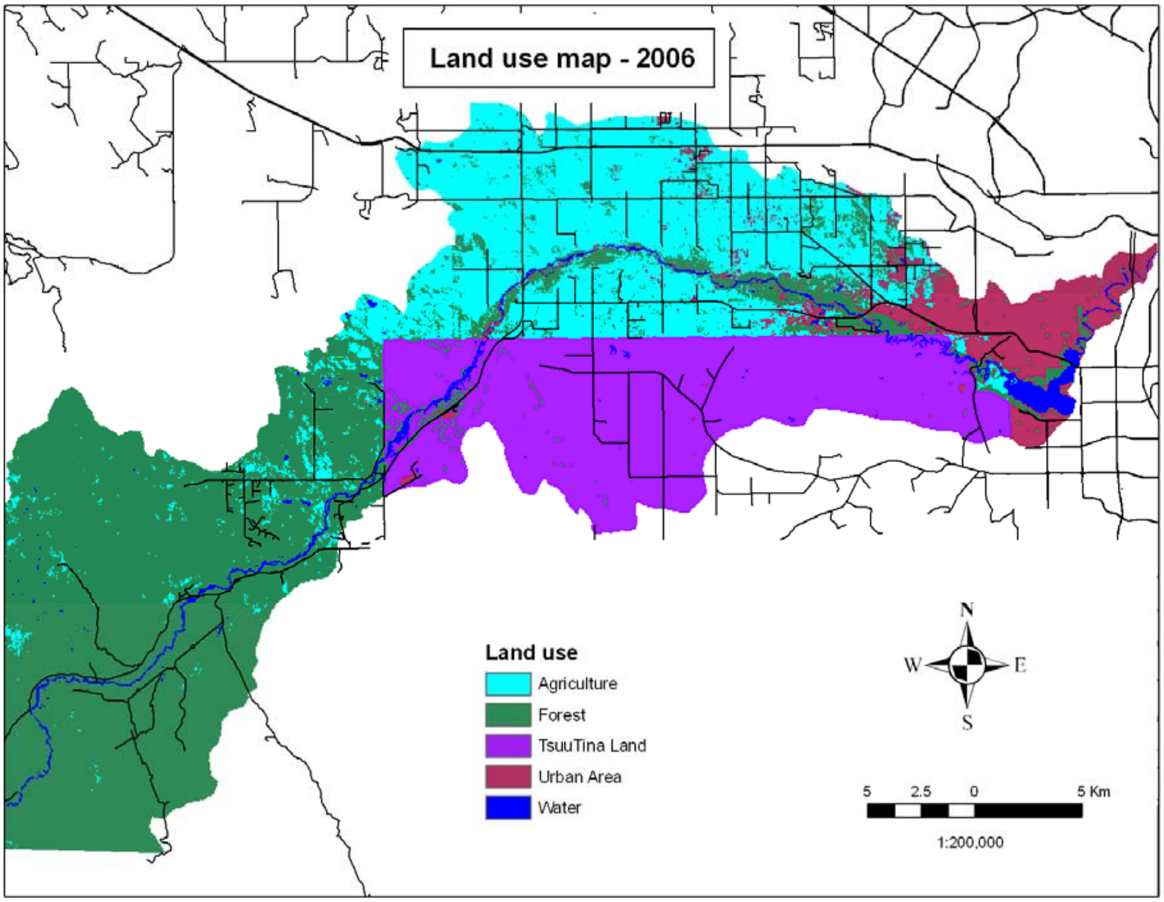

Figure 2.8. Land use map of the Elbow River Watershed – 2006

Table 2.2: Percentage of the study area covered by different land uses in the original land-use maps.

	1985	1992	1996	2001	2006
Agriculture	27.51	26.85	26.44	23.88	25.39
Forest	45.56	45.90	46.08	47.65	46.42
Urban	2.50	3.06	3.42	4.02	4.80

Driving factors were considered in the transition rules. They are represented by raster images of the same resolution and extent as the land-use maps. The model is designed to accept any number of driving factors, though only three were used in this study:

- distance to downtown Calgary,
- distance to a main road, and

- distance to a main river.

These driving factors are commonly quoted in the literature as important factors influencing land-use changes (Clarke et al., 1997; Messina et al., 1999; Ward et al., 1999; White and Engelen, 2000; Candau, 2002; Cheng and Masser, 2004; Caruso et al., 2005; Fang et al., 2005; Lau and Kam, 2005; Dietzel and Clarke, 2006a; Benenson, 2007). Moreover, a visual analysis of the location of the historical land-use changes within the study area confirmed that these driving factors were of importance.

The aforementioned distances were calculated for every cell and for each historical year in the study area using the Euclidian distance tool available in ArcGIS 9.1. Downtown Calgary was digitized as a point. The water network available from Ensight Info was used to compute the distance. The road network vector file that was used is also available from Ensight Info. It is probably the most accurate road network representation freely available to researchers in Alberta. The distances were computed using the 2006 data. If available, a user can compute these distances using different data sets such as the road network for each historical year. In such case, the user would have to use different driving factor maps for each simulated time step.

2.2 Model architecture

The CA model is composed of a hierarchy of modules. At first, the CA is divided into two main modules, one for the extraction of the transition rules and for the model calibration, and another one for the simulation. This allows a user to run only one of the modules several times if needed. For instance, it might be convenient to use the same transition rules but to run simulations with different scenarios. Also, if two different study areas share similar characteristics, the transition rules from one study area can be extracted and applied in the other study area.

2.2.1 Preparatory steps to the rule extraction and calibration module

Before extracting the rules and calibrating the model, several preparatory steps must be performed. They include:

- the definition of the neighborhood,
- the procedure implemented to handle the edge effect,
- the proper organization in the computer memory of the land-use maps, the number of cells of each state in the neighborhood and the driving factors, and
- the procedure developed to maximize the use of available memory and reduce computation time.

These steps are described in the following sections.

2.2.1.1 Neighborhood definition

In this model, the neighborhood was designed to approximate a circle around the center cell. The coordinates of the cells forming the neighborhood are computed only once, considering that the center cell is located on position 0,0. These coordinates can be added to the coordinates of any cell in the study area to obtain the coordinates of the cells composing its neighborhood.

Unlike most CA models, in this study, there is no distance function applied on the neighborhood, that is, a cell near or far away from the center cell has the same influence in the transition rules. Instead, to reduce the induced bias from distant cells, several concentric neighborhood rings were defined (Figure 2.9). The different rings are all exclusive, that is, a cell can only be located in a single ring, and there is no gap between two rings. The user can choose the number and size of the rings. Within each ring, the influence of the neighboring cells on the central cell is constant but this influence is different between rings. Consequently, the influence continuous distance function used in most CA model is here a discrete distance function. This approach has the main advantage of greatly simplifying the definition of the cell 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 user to use historical data at any scale without changing the model. This approach is also required by the rule extraction method.

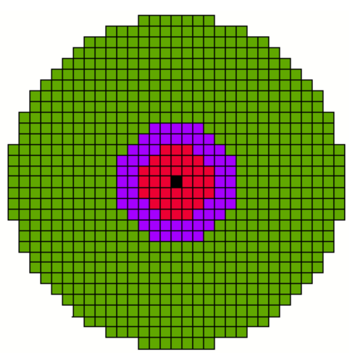

Figure 2.9. Definition of the neighborhood of a central cell. This figure shows three concentric and mutually exclusive rings of three, five and fifteen cells of radius, respectively.

The neighborhood size and configuration have an impact on the outcomes of the model (Liu and Phinn, 2001, Chen and Mynett, 2003, Ménard and Marceau, 2005, Benenson, 2007, Georgoudas *et al.*, 2007) and should therefore not be arbitrarily chosen. If the neighborhood does not correspond to a zone of influence related to a land-use change, the simulation output would be very poor. While testing all the neighborhood configurations is beyond the scope of this study, a sensitivity analysis of the neighborhood sizes composed of 106 different sizes was performed. Neighborhood sizes having a maximum of fifteen cells of radius – or 900 m at 60 m resolution – have been tested. Such maximal distance is greater than what is used in most CA models. These neighborhoods were composed of one, two or three rings. For each tested neighborhood configuration, the model was calibrated and land-use change simulations were performed; the results were compared with known land-use maps, as explained in section 2.4.2. This analysis demonstrated that different neighborhoods led to very different simulation outputs, confirming the sensitivity of the model to the neighborhood size. Consequently, the neighborhood configuration that led to the simulation result that was the most similar with the known land-use maps was used in this study. This neighborhood is composed of three rings, having respectively a radius of three, five and fifteen cells of radius.

2.2.1.2 Edge effect avoidance

CA models use raster images which typically cover a rectangular area. However, it is often of interest to apply the model on a study area of any shape, like an administrative region or a watershed. As a consequence, an important number of cells of the rectangular grid can be tagged as the background which represents all the cells that are outside the study area. Cells located at the edge of the study area have some background cells within their neighborhood and it would be wrong to consider them in the transition rules. To reduce the edge effect, this model only considers the cells that have a neighborhood including no background cells.

Two main techniques are applied in current CA models to address the edge problem and to simulate state transition on every cell of the study area. The first method is to wrap around the study area so the first and last lines are considered as subsequent, as well as the first and last columns. This technique produces some significant artifacts as important landscape features can be duplicated. One can imagine a city located at the edge of the study area and an agricultural zone on the other end. It is wrong to assume that the agricultural cells are neighboring the city and should therefore be transformed to urban. Also, this method can easily be applied when the study area is a rectangle but it becomes difficult to assess which cell should be connected to which one when the study area is of a more complex shape like a watershed. Should a cell be connected to its upper or lower neighbor, or should it be to its left or right neighbor? For this reason this technique has not been applied in this study.

The second method is to replicate the cells forming the edge until a neighborhood with no background cell exists for every cell of the study area. This technique is appropriate when the landscape is quite homogeneous but brings an important bias when it is not. One can think of a small village located at the edge of the study area, in the middle of an agricultural zone. After duplication, there is not a small village anymore but rather a city which will alter the probability of changing the surrounding cells. Consequently, this method has also been discarded in our study.

To reduce the edge effect to its minimum, this CA model was applied only on the cells in the study area that have a neighborhood with no background cell. A user can use land-use maps that are larger than the study area, or let the program reduce the extent of the study area over which simulation will be conducted. This method is however introducing a bias if the study area ends close to a land-use feature that is very different from the land uses at the edge. For example, if the study area ends at a city limit, the cells forming the edge will remain agricultural or forested while in reality they have a high chance of being urbanized. This bias is however minor if land-use features outside the study area that are very different from the land uses at the edge are small in size; otherwise, they should be considered within the study area. Moreover, the user can always manually update the state of the cells located at the edge of the study area at each simulated time step. During the calibration, there is no bias as the state transitions of the cells in the edge are simply not considered.

2.2.1.3 Layering of the neighborhood characteristics and driving factor values

After the identification of the cells having no background cell within their neighborhood, the model must count, for these cells and for each historical map, the number of cells of each land-use type that are within each neighborhood ring. This information is stored in memory as a set of spatially overlaid layers allowing the retrieval of the neighborhood composition of a given cell by looking at the values in each layer at the spatial location of the cell. For each historical date, the model stores a spatial layer for the land use, one layer for each land-use type in each neighborhood ring and one layer for each driving factor (Figure 2.10).

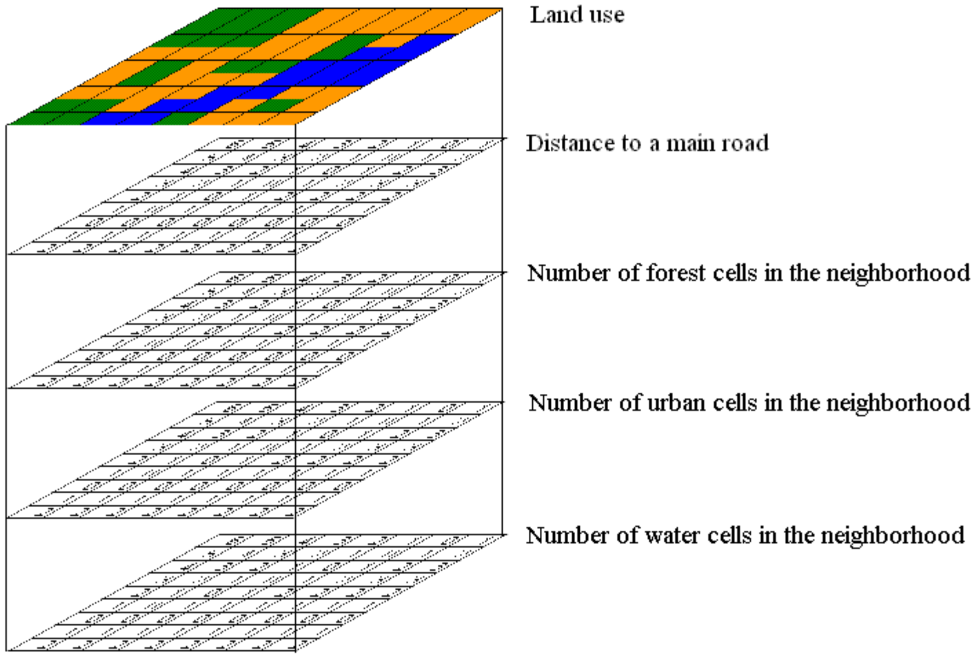

Figure 2.10. Spatial overlay of the information referring to each cell.

The model can read the value of a given cell in each layer; this is very useful because for both the rule extraction module and the simulation module, this information is required several times for each cell. Finding the neighborhood characteristics is therefore done only once, speeding the entire calibration process. This approach is memory consuming for large study areas though it remains reasonable. The maximum required memory for this feature can be computed using Equation 1.

Maximum required memory = number of cells in the study area * 32 bits * (number of land-use classes * number of neighborhood rings + number of driving factors). Equation 1

The amount of required memory is likely smaller as some layers may be of a different data type, like a byte or integer requiring only 8 or 16 bits per cell. As an example, using six land-use classes, three factors and two neighborhood rings in the Elbow River watershed which is about 1 500 000 cells at a 30 m resolution, the maximum required memory is:

1 500 000 cells * 32 bits * (6 classes * 2 rings + 3 factors) = 85 Mb.

Since the required memory can be important, this CA keeps only the cells forming the study area and discards the background cells using a 2D to 1D map transformation.

2.2.1.4 Use of pseudo 1D maps

For the Elbow River watershed, at 30 m resolution, the complete grid is composed of 4 500 000 cells while the study area accounts for 1 500 000 cells, or 1/3 of the grid, the remaining ones being the background. To minimize the memory load, 2D grids were transformed into 1D arrays. Row after row, the non-background values were copied from left to right and were added to the end of a 1D array as illustrated in Figure 2.11. To retrieve a cell in the 1D array from the original grid, an index grid covering the whole rectangular grid was created, representing a flag value over the background cells and the position index in the 1D array for each cell of the study area. Similarly, another array of the same size as the 1D array held the position of each cell in the original grid.

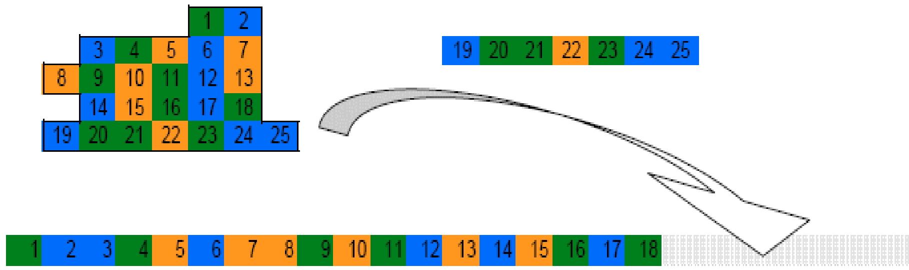

Figure 2.11. 2D to 1D map transformation. Each row of data is copied at the end of a 1D array to remove all background cells.

Except for the index array, each layer was transformed to a 1D array, considerably saving memory. The required memory to handle the data in the Elbow River watershed was about 85 Mb using this technique compared to 255 Mb without any 2D to 1D transformation.

The model now has access to all the information that is required for extracting the transition rules. The concept and detail of the rule extraction and calibration module are presented in the next section.

2.2.2 Rule extraction engine and model calibration

The calibration method developed in this study is different from previous CA models in respect to several aspects:

- there is no limitation on the number or type of driving factors and land-use classes;
- multiple transition rules describe a land-use change. These rules are dynamically found by analyzing historical data. Instead of having a different transition rule for

each possible combination of individual values in each layer, the rules are defined for ranges of values in each layer;

- graphical and interactive displays allow the user to find the level of influence of a driving factor or a particular land use and to define conditional rules;
- simple statistics are computed by the CA model to transform the conditional rules into mathematical rules, which are a representation of the driving factor values and neighborhood configuration. During the simulation, for each cell of the study area, the similarity between the rules and the driving factors values and neighborhood configuration is assessed to establish which rule should be applied;
- the influence of each rule at each historical date is computed so temporal trends can be captured and simulated.

The rule extraction and calibration module includes five main steps:

- read the set of historical land-use maps and driving factors; these maps must be supplied by the user;
- recursively, for each type of land-use change, select all the cells that have changed state in all the historical land-use maps;
- display frequency histograms plotting the value of the driving factors and the number of cells of each land use located within the neighborhood of the selected cells;
- the user interactively selects the ranges of values based on the interpretation of the histograms;

- the model combines these ranges of values for each land-use type and for each driving factor and transforms them into parameters representing the transition rules.

The user must specify which land-use changes should be considered as some land-use changes may be of little interest to simulate. In this study, the model was calibrated for four types of land-use changes:

- from forest to agriculture,
- from forest to urban,
- from agriculture to forest,
- from agriculture to urban.

Other changes have not been considered for different reasons. Changes from and to Tsuu Tina land were the result of improper classification as the boundaries of this territory did not change between 1985 and 2006. Changes from and to water were mainly a result of topographic factors, hydrological cycles and intensity of precipitation and this information was not considered in this model. At each historical date, between 28% and 42% of the water cells changed to another land use. Not simulating this change introduced discrepancies between the model outcomes and the reality, especially if other land-use transitions relied on the number of water cells. Changes from urban to any other land use were marginal in this study area and have not been considered in the simulation. Also, “changes” that kept a cell in its state, such as from Forest to Forest, have not been incorporated. A cell may have a greater potential of staying in its current state than changing to another state. Preliminary tests have shown that simulations with and without these “changes” produce similar results but considerably reduce the execution time.

Using the user-specified neighborhood rings, the model creates all the layers corresponding to the number of cells of each state in each neighborhood ring. In this study, the neighborhood is composed of three rings having a radius of three, five and fifteen cells, corresponding respectively to 180, 300 and 900 meters. This selection is based on the sensitivity analysis described in section 2.2.1.1.

For each user-specified land-use change, the model selects the cells in all the historical maps that have completed this land-use transition. The values of these cells in all layers are displayed one after another in the form of a frequency histogram. By analyzing these histograms, the user can find which values are or are not associated with the given land-use change. The histograms are composed of up to four main basic distributions as shown in Table 2.2. Of course, the distribution in the histogram can be – and is often – a combination of these elementary distributions. If a particular or a range of layer values are associated with a land-use change, these values will often be present in the layer of the historical data and the histogram will show a peak. On the contrary, if the values are almost never present in the layers of the historical data, then one can assume that these values are associated with no change and the histogram will display a low frequency value. Similarly, if all values in a layer are present the same number of times in the historical data, it means that this value is not associated with the land-use change and a horizontal line will be displayed on the histogram.

Table 2.3: Relation between the shape of the histograms and land-use change

Influence of a particular layer value over a land-use change	Distribution of the histogram
Associated	Normal; modal or multimodal; peak
Not associated	Uniform
Associated with no change	Low occurrence
No obvious influence	No particular distribution

The task of the user is therefore to identify these shapes on the histograms, especially the peaks of values associated with a land-use change, by selecting one or more ranges, also called groups, of values on the histogram. This step could be automated, though the user learns a lot about the processes behind the land-use changes by manually selecting the values. The user must look for ranges of contiguous values having a normal distribution, or as close as such a distribution as possible. With such a distribution, the closer a layer value is to the mode, the better is the association with the land-use change, and during the simulation, the better are the chances of having this land-use change. Inflection points in the cumulative distribution curve provide an additional help to the user to identify such ranges of values. For example, Figure 2.12 shows an almost normal distribution centered on 14 forested cells. During the simulation, the closer the number of forested cells to 14, the more likely the change is going to happen. The user should identify one group in this histogram, from 0 to 36 cells.

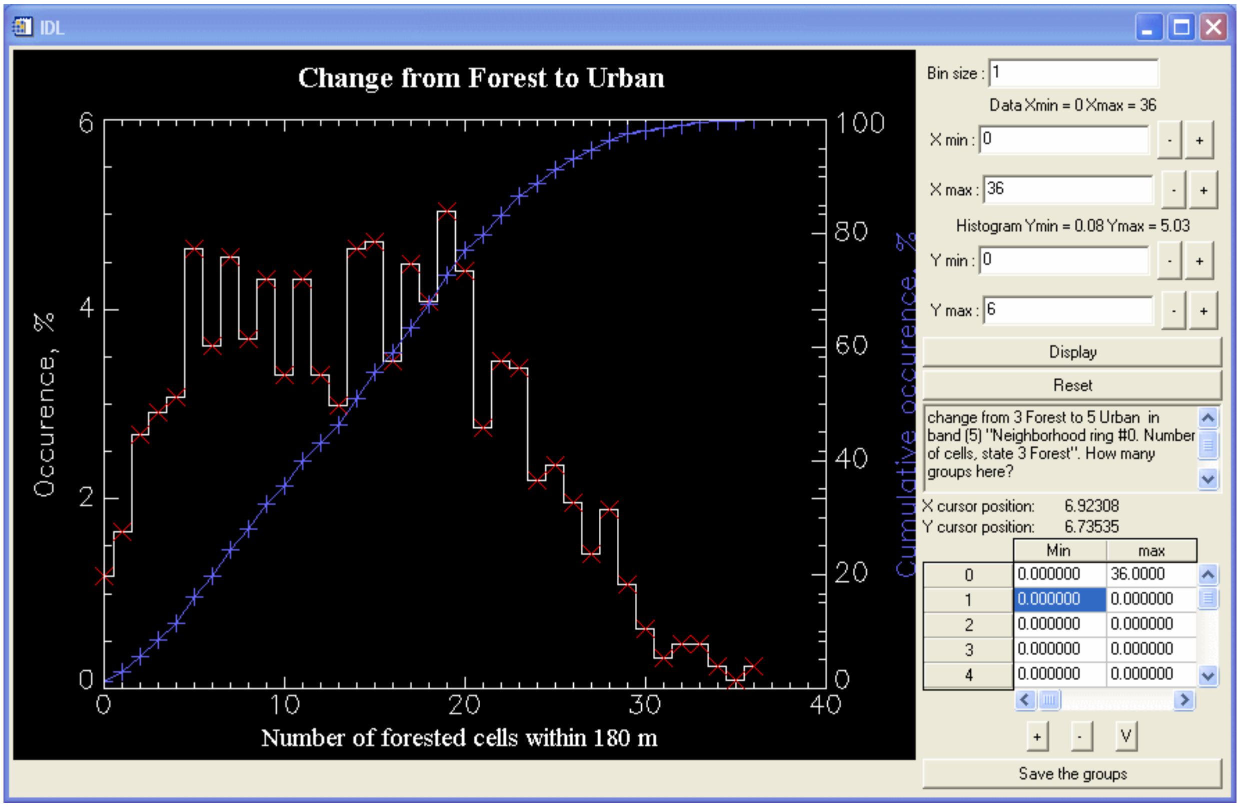

Figure 2.12. Frequency histogram of the number of forested cells located within the first neighborhood ring (0 to 180 m) of the cells that have changed from forest to urban in 1985, 1992, 1996 and 2001. The blue curve represents the cumulative occurrence of all the cells.

It is common to find a particular value associated with a land-use change. In this case, a peak is displayed and the adjacent layer values have a much lower frequency. In such cases, the user must identify this peak value in one "group", and the other values must be assigned to one or more other groups, depending on their distribution. For example, Figure 2.13 shows a peak at zero urban cells in the first neighborhood ring, corresponding to a high probability of change from forest to urban. Then, the distribution is almost uniform from 1 to 36 urban cells, meaning that the probability of change is not be influenced by the number of urban cells. The user should select two or three groups on this histogram, the first one being the value 0, then from 1 to 8 cells, and finally from 9 to 36 cells.

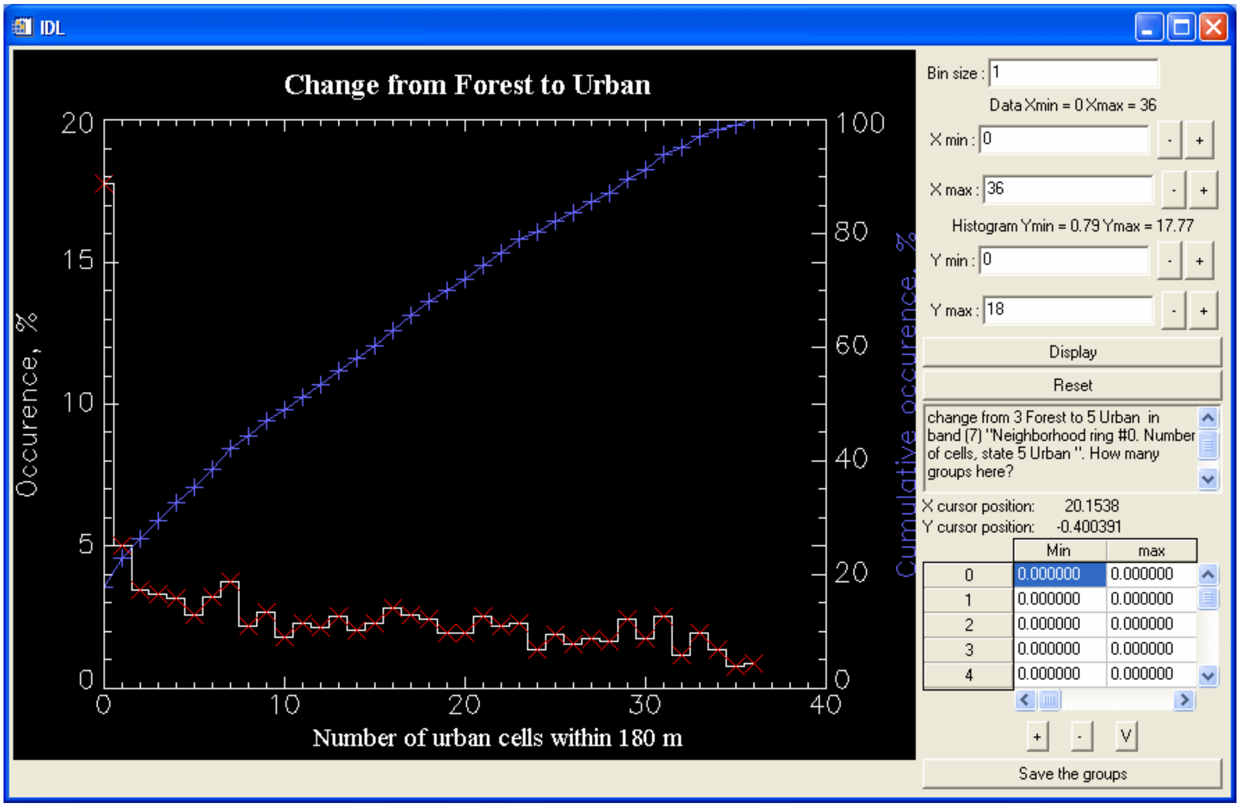

Figure 2.13. Frequency histogram of the number of urban cells located within the first neighborhood ring (0 to 180 m) of the cells that have changed from forest to urban in 1985, 1992, 1996 and 2001. The blue curve represents the cumulative occurrence of all the cells.

A histogram could display multiple groups having a pseudo normal distribution. In this case, the user must put each range of values corresponding to a pseudo normal distribution into different groups. For example, Figure 2.14 shows three groups. There is a first group, from 0 to 50 forested cells highly associated with a land-use change from agriculture to forest. Then, the association becomes weaker as it is less frequent in the historical data, but the change is still possible where there are between 50 and 250 cells. Around 400 cells, a very low influence can be observed, meaning that this land-use change almost never happened when there were around 400 cells of forest within the neighborhood. At last, there is another association to the land-use change when the value in this layer is between

500 and 652 cells. The user should select two groups, one from 50 to 400 cells and another one from 401 to 652 cells. The low values are considered in the groups but since these values are far from the mode of these two groups, the CA model interprets that they usually should prevent a land-use change from agriculture to forest.

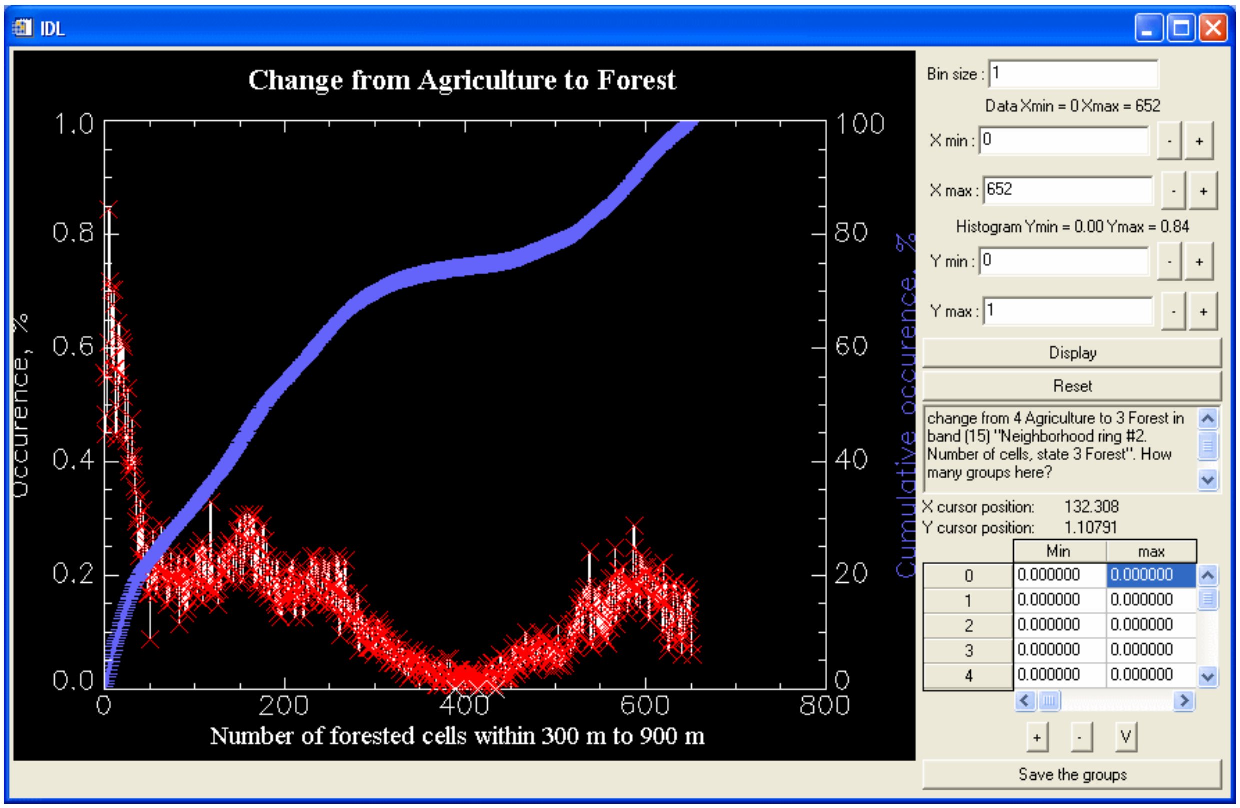

Figure 2.14. Frequency histogram of the number of forested cells located within the third neighborhood ring (300 to 900 m) of the cells that have changed from agriculture to forest in 1985, 1992, 1996 and 2001. The blue curve represents the cumulative occurrence of all the cells. Due to a high number of values on the x axis, the density of points is higher and therefore the look of the curve is altered.

To ensure that every process driving a land-use change will be simulated, the user must set groups that encompass all the possible values. However, it was empirically found that having groups including approximately 90% of the cells provided similar results but drastically increased the speed of the model. The cumulative occurrence curve was used to estimate this percentage. Also, except for peak values, the identified ranges limits can be slightly varied without altering the simulation outputs.

Once the groups have been defined, every group value in each layer is combined with every group value in other layers (Figure 2.15). These group combinations represent the possible neighborhood configurations that are responsible for a land-use change. Most of the combinations are dismissed as they are not present in the historical data. When the cells corresponding to a group combination are selected, the model computes in each layer the mean and the standard deviation of the layer value under the selected cells, that is, of the driving factors and the number of cells of every land use in each neighborhood ring. If a ‘group’ is made of a single value, the mean is this value and the standard deviation is zero. These two values are the parameters of each transition rule.

There are usually many transition rules describing each type of land-use change. The number of transition rules depends on the historical data that are used. It is valid to describe each transition rule by its mean and standard deviation value in each layer as, for all or for some layers, different types of land-use changes and different transition rules have very different values, which is consistent with Verburg *et al.*’s statement (2004): “*different land use types have clearly distinct neighborhood characteristics*”. To visualize this, it is possible to plot, for different types of land-use changes, the value in one layer on one axis and the value of another layer on the other axis (Figure 2.16). Since the different land-use changes have distinct patches, one can say that different land-use changes, and similarly different rules, can be uniquely characterized by their mean and standard deviation value in each layer.

Identified groups of values of interest in each layer:

Cell state	Number of cells in the neighborhood		
	Urban	Forest	Agriculture
urban	1-10	0	0
	11-20	1-100	1-10
			11-20

Corresponding group combinations:

Cell state	Number of cells in the neighborhood		
	Urban	Forest	Agriculture
urban	1-10	0	0
urban	1-10	0	1-10
urban	1-10	0	11-20
urban	1-10	1-100	0
urban	1-10	1-100	1-10
urban	1-10	1-100	11-20
urban	11-20	0	0
urban	11-20	0	1-10
urban	11-20	0	11-20
urban	11-20	1-100	0
urban	11-20	1-100	1-10
urban	11-20	1-100	11-20

Figure 2.15. Correspondence between the groups identified by the user in each layer and the group combinations

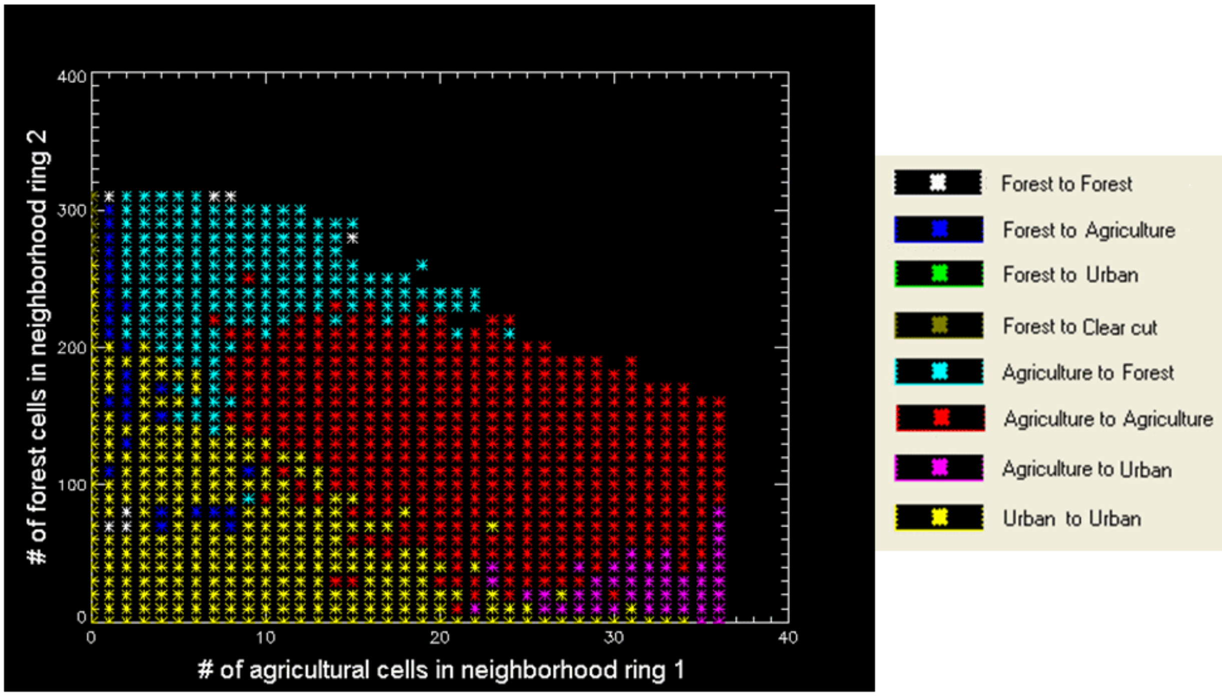

Figure 2.16. Data space of different types of land-use changes. In this figure, the number of agricultural cells in the first neighborhood ring is plotted against the number of forested cells in the second neighborhood ring, for different types of land-use changes.

2.2.2.1 Representation of the transition rules

The model internally generates the transition rules and creates a file providing to the user, for each historical year, the percentage of cells that adhered to each transition rule, as well as the minimum, maximum, mean and standard deviation values for each layer. An example of a transition rule is presented in Table 2.3. This rule corresponds to 0% of the changes for the period 1985 – 1992, 0.4% for 1992 – 1996, 3% for 1996 – 2001, and 0% for 2001 – 2006. When looking at all the possible transition rules, one can note that for the changes from Forest and Agriculture to Urban, most transition rules are historically applied at one date only, while the rules driving the changes from Forest to Agriculture and from Agriculture to Forest are mainly applied at all historical dates. Moreover, the number of transition rules describing each land-use transition is very uneven, as we can see in Table 2.4. Simulation using either all the transition rules, disregarding transition rules making only one cell change its state in all the historical data, or disregarding all the rules that had less than 0.1% of influence led to very similar results. Evidently, the smaller the number of transition rules there are, the faster the execution of the simulation. This resemblance in results is caused by the fact that some transition rules are probably very similar to each other and therefore removing the rules that have little influence causes an increase in the influence of the remaining rules that corresponds to a very similar process driving land-use change as the influence of all the rules is always scaled to 100%. Consequently, the model can either use all the transition rules or can only keep the most important ones. The number of transition rules is much greater than what is usually programmed into CA models but remains at a manageable level.

Table 2.4: Example of a transition rule. Description of one of the most important rules for land-use change from Forest to Urban, using data at 60 m resolution for 1985, 1992, 1996, 2001 and 2006 for the calibration, as well as the distance to a main road, to downtown Calgary and to a main river.

		Min	Max	Mean	StdDev
	Cell state	3	3	3	0
	Dist. to main road	115.485	471.154	244.428	88.5125
	Dist. to city center	1269.33	8769.23	3945.33	183.72
	Dist. to river	0	796.154	477.345	43.3187
Neighb. Ring #0 (180m)	Nb cells state 1 Water	0	0	0	0
	Nb cells state 3 Forest	0	36	8	2.50713
	Nb cells state 4 Agriculture	1	15	10	3.4641
	Nb cells state 5 Urban	10	36	18	3.38062
	Nb cells state 6 Tsuu T'ina Nation Land	0	0	0	0
Neighb. Ring #1 (180m -> 300m)	Nb cells state 1 Water	0	0	0	0
	Nb cells state 3 Forest	0	35	18.625	4.03334
	Nb cells state 4 Agriculture	1	31	12.875	4.61171
	Nb cells state 5 Urban	3	38	28.5	3.96412
	Nb cells state 6 Tsuu T'ina Nation Land	0	0	0	0
Neighb. Ring #2 (300m -> 900m)	Nb cells state 1 Water	23	80	46.125	10.8026
	Nb cells state 3 Forest	106	484	194.375	31.955
	Nb cells state 4 Agriculture	112	303	221.125	29.6476
	Nb cells state 5 Urban	146	449	190.375	14.9182
	Nb cells state 6 Tsuu T'ina Nation Land	0	0	0	0

Table 2.5: Number of rules describing each land-use transition.

	Total number of transition rules	Number of transition rules without transition rules making only one cell change its state in all the historical data	Number of transition rules above the 0.1% threshold
Forest → Agriculture	3 316	1 760	781
Forest → Urban	684	231	231
Agriculture → Forest	4 315	2 212	771
Agriculture → Urban	1 270	311	311

2.2.3 Simulation module

The simulation module includes three main steps:

- read the transition rules established in the previous module, as well as a land-use map corresponding to the initial conditions of the simulation;
- for each time step, the model looks at the neighborhood configuration of every cell and computes the level of correspondence with the parameters of the transition rules;
- with respect to the user-specified constraints and to the influence of each rule, the cells change state according to the rule having the highest level of correspondence.

The time step of the simulation is chosen by the user who entered the date at which he/she wants an output to be created. However, the length of the time step should be the same as the time interval present between each historical land-use maps. Different processes change the landscape at different speeds as illustrated by the establishment of a house versus the establishment of a community, and it would be wrong to capture the transition rules at one temporal scale and to simulate land-use changes at another temporal scale.

The first step of the simulation is to load the transition rules and the initial conditions which are the land-use map and the driving factors. For each simulated time step, the model must re-create all the layers and if specified by the user, load new driving factor maps. Then, for each cell, the resemblance between the driving factor values and the neighborhood configuration that led to a transition rule and the driving factor values and the neighborhood configuration of the cell is computed. If they are very similar, it is likely that the cell will adhere to the corresponding type of land-use change. A Resemblance Index (RI) (inspired by the Minimum Distance to Class Mean remote-sensing image classification algorithm (Richards, 2006)) is computed for every transition rule using Equation 2.

$$RI = \sum_{i=1}^{m} \frac{\left| n_i - \bar{x}_i \right|}{\sigma_i} \qquad \text{Equation 2}$$

where *m* is the number of layers, n_i is the value in layer *i*, $\bar{x}_i$ is the mean value for layer *i* in the rule and σ_i is the standard deviation for layer *i* in the 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, as defined in IDL, the programming language used to code this model. Accordingly, $RI \in \Re^{+}$ and the smaller RI is, the more similar is the cell neighborhood configuration to the transition rule.

2.2.3.1 Influence of each rule

Knowing RI for every rule for every type of land-use change is not enough to decide which cell should change its state to another state. The same neighborhood configuration might lead to different land-use changes, and because of some constraints, it might lead to no state change. Therefore, the model must know the influence of each rule. When the transition rule parameters are computed, the percentage of cells that adhere to each type of land-use change, as well as the percentage of cells that adhere to each transition rule, at each year corresponding to the historical data, are computed and saved. During the simulation, the model extrapolates for each simulated year the number of cells that must adhere to each type of land-use change by averaging the number of cells that adhered to each type of land-use change in the historical data and in the already simulated years. The sum for all types of land-use changes of these numbers of cells is scaled down, if needed, so it does not exceed the number of cells on the initial condition map that has a land use implicated in any type of land-use transition. In other words, if some land uses are not considered in any type of land-use change, like the Tsuu Tina class, then the number of cells of this class is deducted from the possible number of cells that can change. The number of cells to adhere to each type of land-use change can also be submitted by the user as a constraint, as we will see in section 2.3. Similarly, the number of cells that must adhere to each rule is extrapolated for each simulated date and scaled to 100%, using the average of the historical and already computed influences.

2.2.3.2 Assignation of a rule to each cell and for each type of land-use change

The selection of the cells that must adhere to each rule is done in a recursive way, for each type of land-use change. The model selects all the cells that could adhere to the first rule, and keeps the best cells, the ones with the lowest RI value, until the percentage of cells that should adhere to this rule is reached. Then the model moves to the next rule, computes the total number of cells that has been selected for all rules so far and retains the best cells until the correct amount is reached. If a cell has a lower RI value than the RI value computed for another rule, the lowest value is kept. Once every transition rule has been considered, the model locks the cells that have been assigned to a rule and re-starts selecting and assigning the remaining cells for the first rule and so on. When no more cells can be selected or assigned to any transition rule, the model moves to the next type of land-use change and repeats the operation. This procedure creates two layers for each type of land-use change: one holding the RI value and one holding the transition rule ID that has the corresponding RI value.

2.2.3.3 Selection of the land-use change to be applied to each cell

We now have, for every cell, an RI value for each type of land-use change. Recursively, the model randomly sorts the type of land-use changes and selects the best cell, i.e. the cell having the smallest RI value, for each of them. If multiple cells are the best ones, a random choice is made. Once the required number of cells that should adhere 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 which are the percentage of cells that adheres to each rule and each type of change. Moreover, a correction factor is computed by deducting the targeted percentages of cells that adhered to each rule and each type of land-use change from the real percentage of cells that adhered to the rules and the type of changes. This information is saved and the model moves to the next iteration. At the next time step, after extrapolating the new statistics as explained in section 2.2.3.1, the model applies the correction so if a rule is over-represented during the previous iteration, it will have less influence at the next iteration, and vice-versa. This feature is interesting as it brakes the linearity imposed by the extrapolation method and therefore better mimics land-use changes.

2.3 Constraints

In this model, a constraint is an external condition that must be satisfied. Three types of constraints were implemented at two spatial scales:

1. No development zone. Over one or more zones defined by polygons, one or more types of land-use change are forbidden. For example, there can not be any urbanization within a provincial park.
2. For the whole study area, a targeted number of cells that should adhere to one or more types of land-use change are specified by the user. For example, the population of Calgary is expected to grow by 2.3% in 2008 (City of Calgary, 2006).
3. For the whole study area, one or more types of land-use change are encouraged or restrained. The chances of development of every cell, for the specified land-use changes, are increased or decreased. For example, the government may decide to

subsidize farmers who keep trees on their land. The chances of a change from forest to agriculture are therefore decreased by a certain percentage.

The constraints 2 and 3 can also be applied at a local level, in which case the user specifies the constraints within one or more polygons. For example, one can simulate the growth of a particular town within the study area. These functions are however not fully operational yet.

To ease the definition of the constraints, a key value can be used to constrain multiple types of land-use change at the same time. For example, the chances of having a land-use change from forest to any other land use can be decreased, or a target number of new urban cells can be added without caring about the initial state of the cells. It is also possible to mix these constraints to simulate a variety of “what if” scenarios.

Internally, the model first computes the number of cells that must adhere to each type of change. Then, it reads the global level constraints from a file and replaces the computed numbers by the read ones. During the simulation, the model first tries to satisfy the local level constraints by selecting the cells having, for each type of change specified in the constraint, the best RI within the specified polygons, one at a time, and deducts the number of cells that adhere to each type of land-use change from the total targets. After this, the model locks the cells over the polygons having a local constraint so that the local targets are not exceeded. The lock is done by setting the RI value of these cells, for the specified land-use changes, to positive infinity. Then, the model allocates the cells in the study area to the remaining targets, as explained in section 2.2.3.3.

2.4 Output validation

Validating the model output is of the utmost importance to ensure confidence in the results. Three types of validation have been done in this study. First, to know if the model is able to establish adequate transition rules, a fictional dataset created by known rules has been used in the calibration, and the resulting rules are compared with the original ones. Second, to assess the quality of the calibration, results of a simulation conducted from the past (1985) to the present (2006) were compared against known land-use maps. Then, once confidence in the model has been confirmed, simulations were run from the present (2006) to the future (2031) to forecast land-use changes. In this case, since no known land-use maps exist to validate the results, the opinions of a group of experts were collected.

2.4.1 Validation of the rule extraction method: Conway's game of life

To assess the ability of the model to establish the proper transition rules, maps were created following the Conway's game of life (Gardner, 1970). These maps were supplied to the CA model as "land-use maps" and the calibration was performed. The transition rules found by the CA model were then compared to the original rules.

The process in the Conway's game can be defined as follow. The cell state is either "dead" or "alive", the neighborhood is defined as the eight adjacent cells and four rules dictate the change of states:

- if a cell is dead with exactly three alive neighbors then change the cell state to alive;
- if a cell is dead with more or less than three alive neighbors then the cell remains dead;
- if a cell is alive with 1, 4, 5, 6, 7 or 8 alive neighbors then the cell state changes to dead;
- if a cell is alive and has exactly two or three alive neighbors then the cell remains alive.

Using the rules of the Conway's game of life has several advantages. The number of cell states is as small as possible, reducing the number of possible cell state transitions. The number of rules is also very limited, allowing an easy comparison with the CA model transition rules. Despite these characteristics, the Conway's CA model is nonetheless able to exhibit a complex dynamic.

2.4.2 Validation of the calibration: Simulation from the past to the present

For the second type of validation, two maps must be compared. Several methods exist, each of them having their limitations. The first method is undoubtedly the visual comparison of the simulation result and of a known land-use map. This method is efficient when the results are clearly wrong and for assessing whether the results are plausible or not (Wu, 1998; Yeh and Li, 2001). On the other hand, two maps can look very similar but have very little in common when compared pixel by pixel (Jenerette and Wu, 2001; Liu and Andersson, 2004). This method also relies on the operator knowledge of landscape patterns and judgment abilities (Wu, 2002).

Automated methods have also been proposed. Pixel to pixel comparison is frequently used (White and Engelen, 2000; Wu, 2002; Li and Yeh, 2004). This method is fast, but since CA models are better for modeling patterns rather than microscopic components, it often provides poor results (Yeh and Li, 2005). Hagen (2003) introduced the concept of fuzzy Kappa in which each cell has more than one state as it is 'composed' of the surrounding cells. It is also possible to compare land-use patterns via metrics describing the territory such as fractal dimension (Yeh and Li, 2001; Wu, 2002; Kocabas and Dragicevic, 2004), number of patches (White and Engelen, 2000; Jantz and Goetz, 2005), connectivity between patches (Ménard and Marceau, 2007), polygon overlapping (White, 2006) and object equivalence (Abdelmoty and El-Geresy, 2000). However, patterns being very different can have the same value for one of the index (White and Engelen, 1993; White and Engelen, 2000), and it is frequent to have different indices indicating contradictory level of equivalence between the maps (Straatman *et al.*, 2004). Dietzel and Clarke (2007) describe a way to combine several metrics together, easing the comparison.

Cell to cell comparison is not the best way to compare the maps; rather, patterns should be compared. The later option is however nontrivial to implement. The target of the validation method proposed in this thesis is therefore to do better than a cell to cell comparison, but almost as easy to implement. Instead of comparing a single cell between maps, the idea is to compare the content of the cells within a neighborhood of each cell, dismissing the spatial location of the cells within the neighborhood. In other words, the states of the cells forming the neighborhood of a cell are assigned to this center cell. Then, the comparison is done by computing the percentage of correspondence between the simulated map and a known land-use map, by dividing the number of corresponding correct cell states by the total number of cells within the neighborhood, as illustrated in Figure 2.17. This value is saved at the location of the cell in the center of the neighborhood. This results in a map covering the study area.

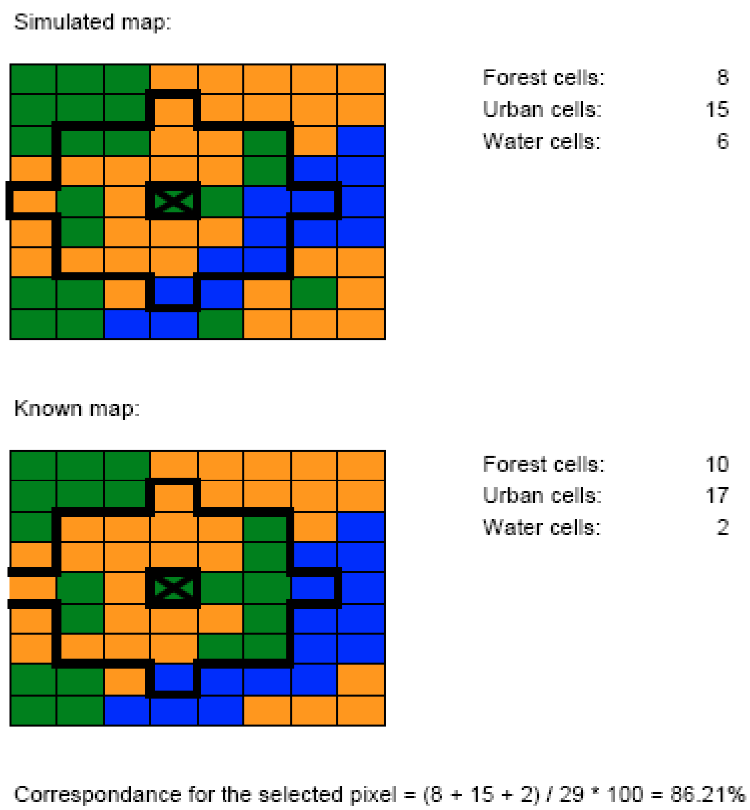

Figure 2.17. Output validation.

With this method, the exact location of a cell is not considered, but rather the content of multiple cells within a specified area is taken into account. The user is free to use a circular neighborhood of any size. If the size is small, the output will be zones of important errors followed by zones of very small errors or free of errors. On the other hand, if the neighborhood size is medium, the output will have less or no zones of large errors, but greater areas will have a small or medium error. At last, if the neighborhood size is very large, the error will be smoothed over the whole map and the output will show almost no error. At a resolution of 60 meters, we empirically found that using a neighborhood size of 5, 6 or 7 cells produced the highest average correspondence. A neighborhood size of 5 cells has been used in this study.

2.4.3 Validation of the simulation: opinion from experts

For the simulation results up to the year 2031, the land-use maps created by the CA model were presented to a group of planners and Calgary Regional Partnership representatives to assess their quality. The group was constituted by a senior planner from the city of Calgary, a planner from the municipal district of Rocky View, a project manager from the CRP, a consultant creating land-use maps of the study area and a scholar from the geography department at the University of Calgary having an ongoing project within the study area. Some of these people have access to detailed land-use plans and were able to compare the location of new urban developments proposed by the CA model with the real location of the planned urban developments. Moreover, they have an expert knowledge about the type and location of the main agriculture fields. At last, these people were able to suggest additional driving factors to be considered in the model, such as a map of the wetlands or of the flood plain.

2.5 Simulations performed

Two main groups of simulations were performed: one from the past (1985) to the present (2006) and a second from the present (2006) to 2031. Such temporal extent was chosen to fulfill the needs of the CRP and because the socio-economic situation within the watershed is assumed to remain fairly constant during this period of time.

2.5.1 From the past (1985) to the present (2006)

To test the quality of the calibration, the model was calibrated using the land-use maps of 1981, 1992, 1996 and 2001 considering three driving factors: the distance to downtown Calgary, the distance to a main river and the distance to a main road. The neighborhood was composed of three rings having respectively a size of three, five and fifteen cells of radius. Then, the model was run starting with the land-use map of 1985 as initial conditions, and producing outputs for the years 1992, 1996, 2001 and 2006. The last output map was compared with the known land-use map of that year.

2.5.2 From the present (2006) to the future (2031)

To predict future land uses, the model was calibrated using the land-use maps of 1981, 1992, 1996, 2001 and 2006, considering the same three driving factors and neighborhood rings as described in the previous section. Simulation results were produced for the years 2011, 2016, 2021, 2026 and 2031, using the land-use map of 2006 as initial conditions.

The transition rules established by the model during the calibration were applied for a total of seven simulations regrouped into three categories of scenarios. The first scenario is called “business as usual” where the model internally predicts the number of cells that should change their state by only taking historical trends into account. First, a simulation was run with the aforementioned condition. Then, to assess the impact of the choice and number of driving factors, simulation results for the year 2031 were produced using only two driving factors, namely the distance to a main road and to downtown Calgary.

In the second scenario, land-use changes are constrained by population growth. The number of required new urban cells is specified to the model for each simulated year. Population growth predictions were established by the company CH2M Hill (a consultant of the CRP), for the years 2010, 2020, 2030 and 2075. These predictions, by towns, were transformed into number of pixels for the years 2011, 2016, 2021, 2026 and 2031. To transform the growth per town into a number of pixels, we looked at the proportion of the population of Calgary and of the M.D. of Rocky View that were living in 2001 within the Elbow River Watershed according to the Canadian census data. This proportion was assumed to be constant in the future. The densities within the portion of these towns included in the study area have also been computed. Using the population growth prediction, the portion assigned to the study area was computed and, for each town, divided by the density, indicating the surface of new urban development. These numbers were then divided by the surface of a cell and summed up in order to obtain, for the whole study area, the number of new required urban cells. First, a simulation is run with the aforementioned condition. Then, to test the benefit of including the 2006 land-use map in the calibration, a second simulation where the population growth is constrained was run from 2001 to 2031 with a five year interval using the transition rules found using the land-use maps from 1985 to 2001.

In the third scenario, a virtual new town is included in the initial condition map. The population growth was used as a constraint and new initial conditions were applied, which include a virtual new town on the 2006 land-use map. In some simulations, the driving factor “distance to downtown Calgary” was replaced by a map called “distance to town centers” representing the closest distance to either downtown Calgary or to the new town. A simulation was done using this modified 2006 land-use map as initial conditions. Then, a second simulation was conducted using the modified 2006 land-use maps and by using the driving factor “distance to town centers”. Finally, a third simulation was performed to demonstrate the importance of the driving factor. This simulation used the original land-use map of 2006 with the “distance to town centers” driving factor.

2.6 Programming environment

The CA model developed in this study was written in Interactive Data Language (IDL) version 6.3 from ITT Visual Information Solutions (ITTVIS, 2007), previously known as Research Systems Inc. (RSI). This choice was made for several reasons. IDL has the advantage of being an array-oriented interpreted language based on optimized C routines. As a consequence, one can make an operation on an array at a speed unreachable by a traditional for-loop going through each element of an array. IDL also offers the advantage of having internal functions dealing with spatial data as well as being linked to ENVI, a remote sensing software for image analysis. Finally, IDL is a multiplatform language. As a consequence, the CA can run under Windows, Linux or on a Macintosh. It has been demonstrated that a multiplatform CA tends to be faster than a machine specific CA (Dietzel and Clarke, 2006b).

The initial development of the CA model was done under Windows XP professional then within the Linux environment under Fedora Core 6 to take advantage of additional memory. The CA model can be run under most operating systems; however the efficiency of the model is different according to the operating system, Windows being the worst and Linux the most efficient one.

The program contains 300 functions and procedures. Dividing the program in such sub-modules offers several advantages. First, it is easy to create functions that are optimized and to update them. Debugging these functions is easier as the inputs can be controlled and the outputs can be compared to known values. Secondly, it greatly increases the execution speed as code nested in a loop is shortened to a single line. Finally, a code written in small modules is easy to understand by any programmer, as the structure and the logic of the program appears clearly, without being hidden in specialized code.

Chapter 3: Results

This chapter is divided into three main sections. The first section, which demonstrates the validity of the rule extraction method, presents the transition rules found by the model when calibrating the model with data generated according to the rules of the Conway's game of life. Then, simulation results when the CA model is run from the past to the present are shown to assess the quality of the calibration of the model. In the third section, simulation results of future land-use maps produced using various "what-if" scenarios are presented.

3.1 Rules for the Conway's game of life

Table 3.1 represents the equivalence between the original rules of the Conway's game of life and the ones created by this CA model. Each line in this table represents a transition rule in the CA model. They have been grouped together to ease the comparison with the original rules. By focusing on the cells with a red border in Table 3.1, one can observe that the CA model always find a minimum and maximum number of alive cells in the neighborhood that is consistent with the original rules, demonstrating the effectiveness of the rule extraction method. However, there are some discrepancies between the two models, mainly for the Alive to Dead transition. This is because the data used to calibrate the CA model did not include all the possible neighborhood configurations and the model has no way of knowing them.

Table 3.1: Relation between the original rules of the Conway's game of life and the rules found by the CA model

Change	Original rules in Conway's game of life CA model	CA model - dynamic rules											
		Cell state				Nb cells state 0.Dead				Nb cells state 1.Alive			
		Min	Max	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max	Mean	StdDev
Dead -> Dead	Cell is dead and nb. alive neighbors ≠ 3	0	0	0	0	8	8	8	0	0	0	0	0
		0	0	0	0	6	7	6.55	0.50	1	2	1.45	0.50
		0	0	0	0	1	4	3.14	0.48	4	7	4.86	0.48
Dead -> Alive	cell is dead and nb. alive neighbors = 3	0	0	0	0	5	5	5	0	3	3	3	0
Alive -> Dead	Cell is alive and nb alive neighbors = 4 to 8	1	1	1	0	2	4	3.05	0.74	4	6	4.95	0.74
	Cell is alive and nb alive neighbors = 0 or 1	1	1	1	0	7	7	7	0	1	1	1	0
Alive -> Alive	Cell is alive and nb alive neighbors = 2 or 3	1	1	1	0	6	6	6	0	2	2	2	0
		1	1	1	0	5	5	5	0	3	3	3	0

These results show that the CA model is able to capture simple transition rules. Moreover, the number of rules established by the CA model is close to the number of rules in the Conway's game of life, indicating that the CA model captures the processes driving a cell change of state fairly well.

3.2 Simulation from the past (1985) to the present (2006)

Simulation results initiated with the 1985 land-use map using neighboring ring sizes of three, five and fifteen are shown in Figure 3.1, Figure 3.2, Figure 3.3 and Figure 3.4 for the years 1992, 1996, 2001 and 2006 respectively. The percentage of the study area covered by the different land uses is presented in Table 3.2.

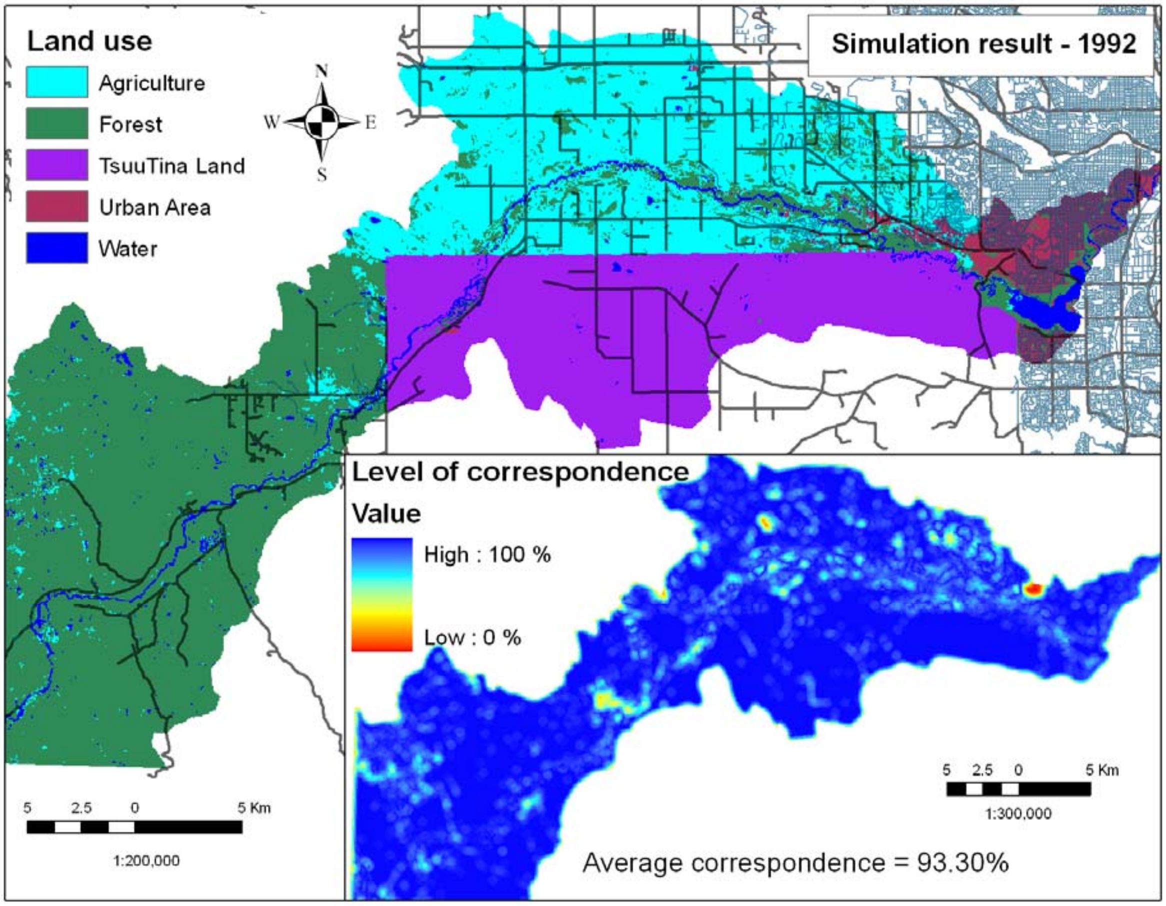

Figure 3.1. Simulation results for the year 1992. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001.

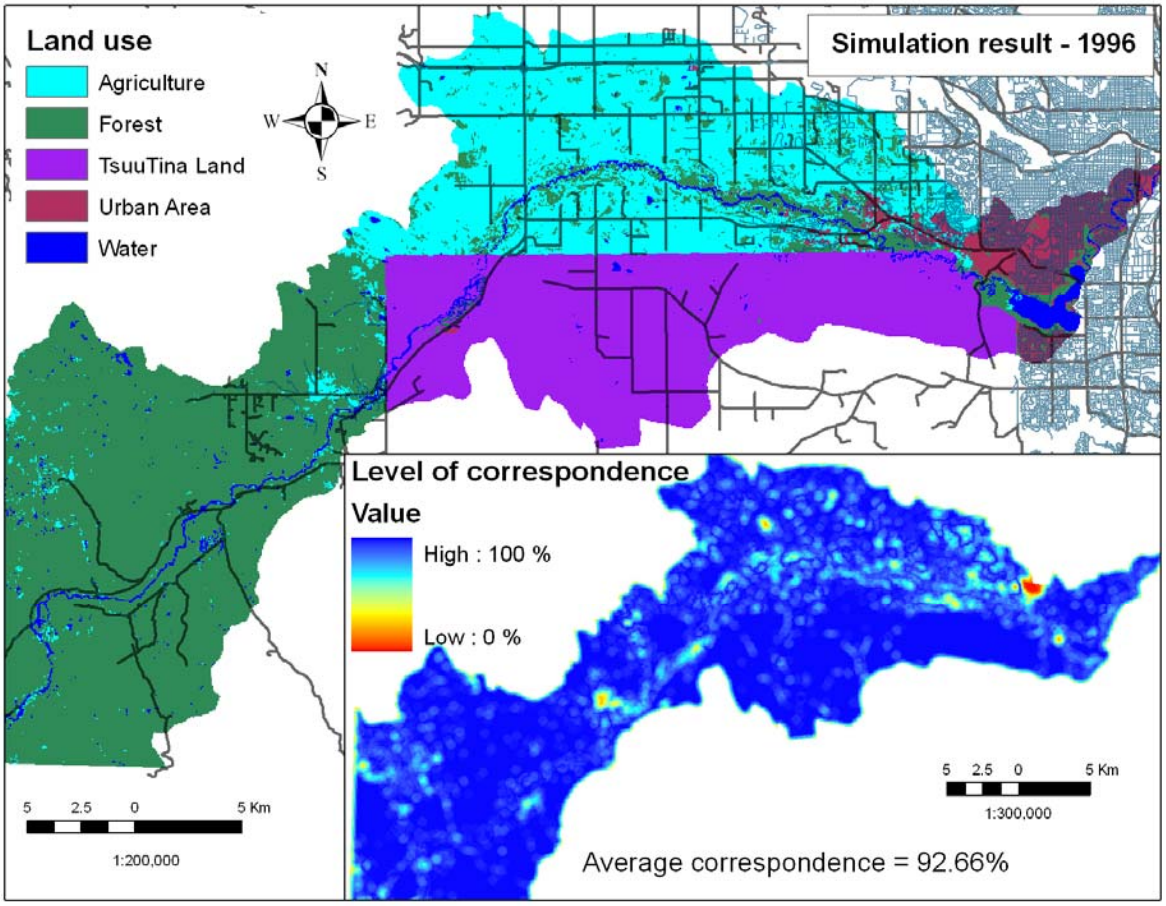

Figure 3.2. Simulation results for the year 1996. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001.

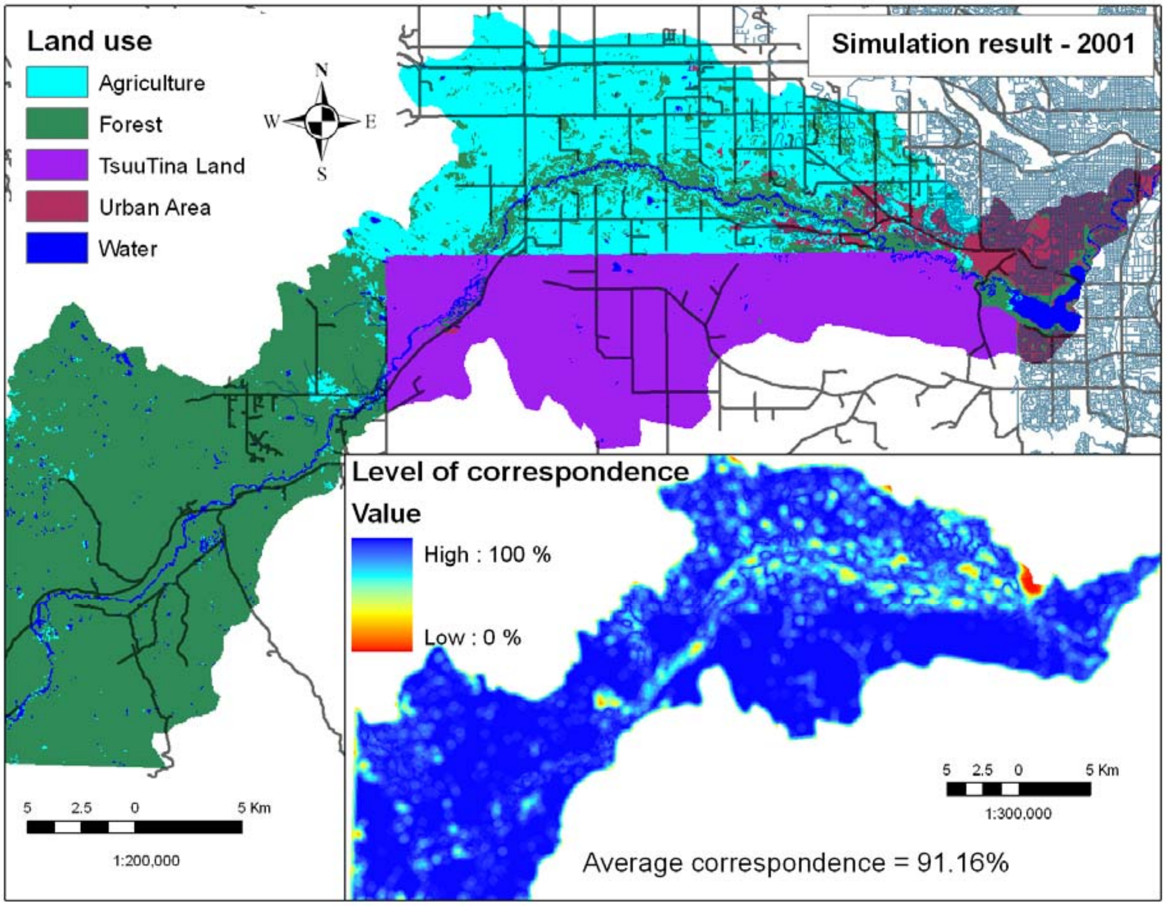

Figure 3.3. Simulation results for the year 2001. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001.

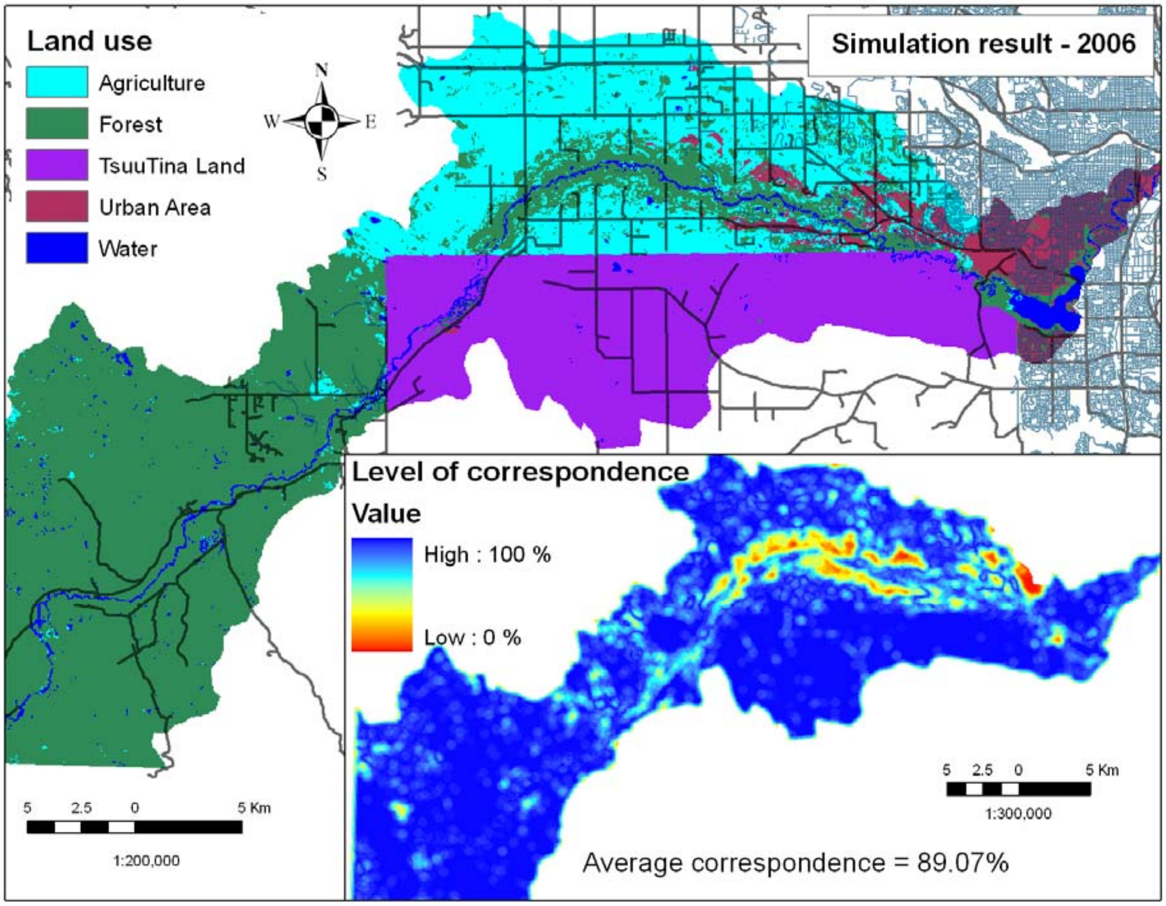

Figure 3.4. Simulation results for the year 2006. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001.

Table 3.2: Percentage of the study area covered by different land uses in the simulation results from the past to the present and variation with the original land-use maps

	1985	1992	1996	2001	2006
Agriculture	27.51	26.87 (+0.03)	26.55 (+0.11)	24.07 (+0.19)	21.28 (-4.11)
Forest	45.56	45.68 (-0.22)	45.63 (-0.45)	47.53 (-0.12)	49.67 (+3.25)
Urban	2.50	3.03 (-0.03)	3.39 (-0.03)	3.98 (-0.04)	4.62 (-0.18)

These results show a level of overall correspondence with the historical land-use maps between 93.3% and 89.07%, which is considerably higher than what is usually stated in the literature (around 70%). When looking at the coverage of each land use over the whole study area, the correspondence between the historical land-use maps and the simulations results varies between 99.97% and 95.89%. The urban extension of Calgary, in the eastern part of the watershed is well simulated. Urban developments are represented by medium to large patches as it was the case up to 2001 in the original land-use maps. These patches are sparsely distributed and located close to main roads and to the Elbow River. This is a realistic scenario for urban development outside of the city limits where people are looking for a nice view and a relaxing environment that is easily accessible. One can observe some inaccuracies between the simulated maps and the original ones mainly along the Elbow River, in the northern part of the watershed. This area corresponds to a zone of very dynamic land-use changes. Moreover, it encompasses the Elbow flood plain. For future simulation, a map of the flood plain might be considered as an additional driving factor.

The errors correspond mainly to the Forest land use. For 2006, the number of forested cells and their location are inaccurate. This can be explained by the fact that greater land-use changes happened between 2001 and 2006 compared to the other temporal intervals, with a large number of cells changing from Forest to Agriculture (Figure 3.5). This information is not available to the model as it has been calibrated with the historical land-use maps up to 2001 only. Finally, since the model does not simulate land-use changes at the edge of the study area, there are additional discrepancies with the historical land-use maps at the limit of Calgary's built up area as these cells remains in the state of 1985.

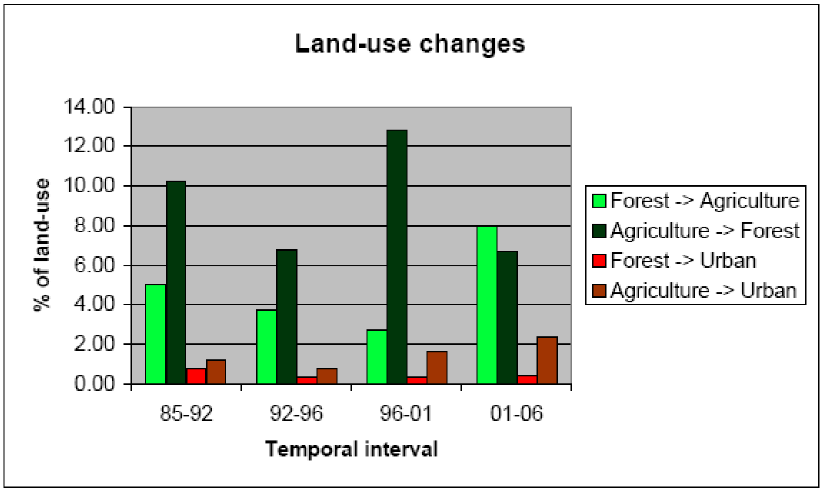

Figure 3.5. Historical rates of land use changes in the Elbow River Watershed

3.3 Simulation from the present (2006) to 2031

This section presents three scenarios of increasing complexity.

3.3.1 Scenario "business as usual"

Simulation results for the years 2011, 2016, 2021, 2026, and 2031 can be seen on Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9 and Figure 3.10 respectively. A close-up showing the western limit of Calgary is also provided, as it is the area where most changes happen. The percentage of the study area covered by the different land uses is presented in Table 3.3.

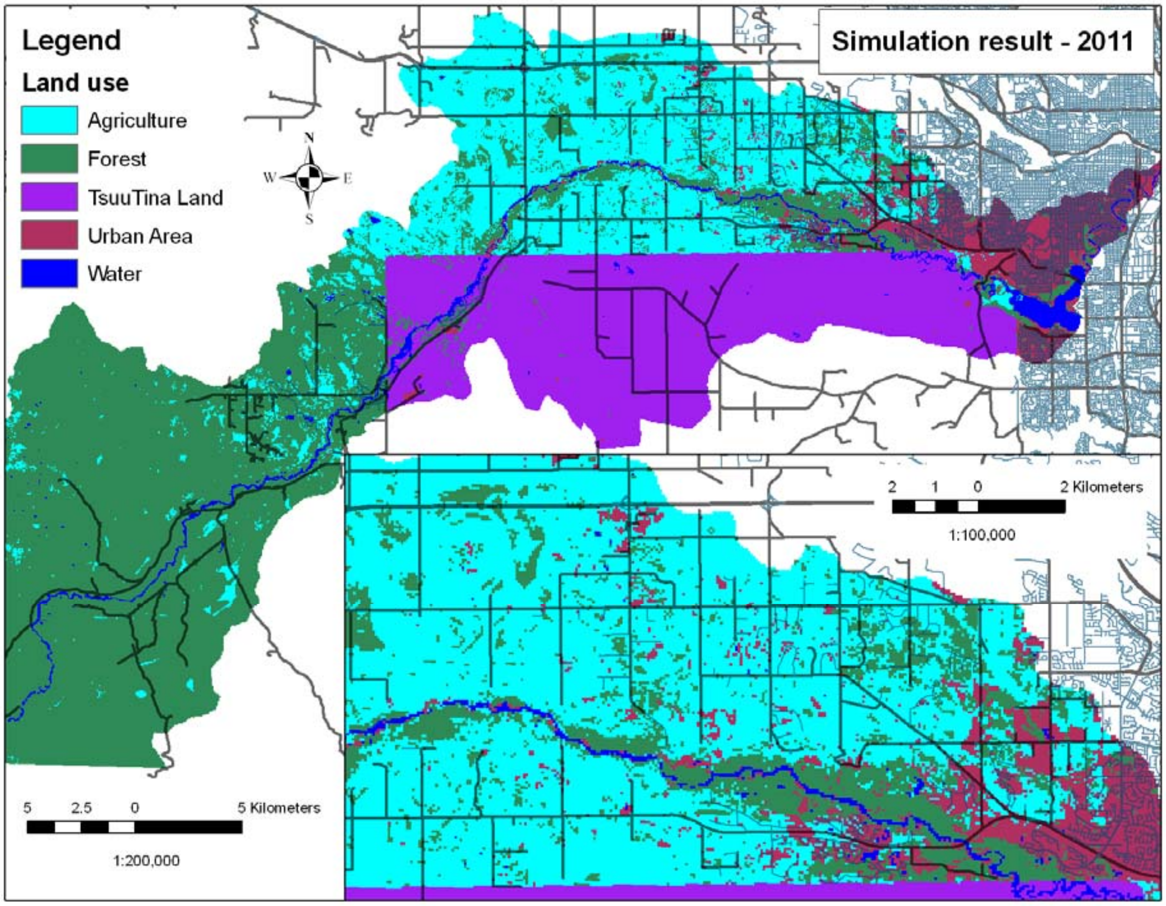

Figure 3.6. Simulation results for the year 2011. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006.

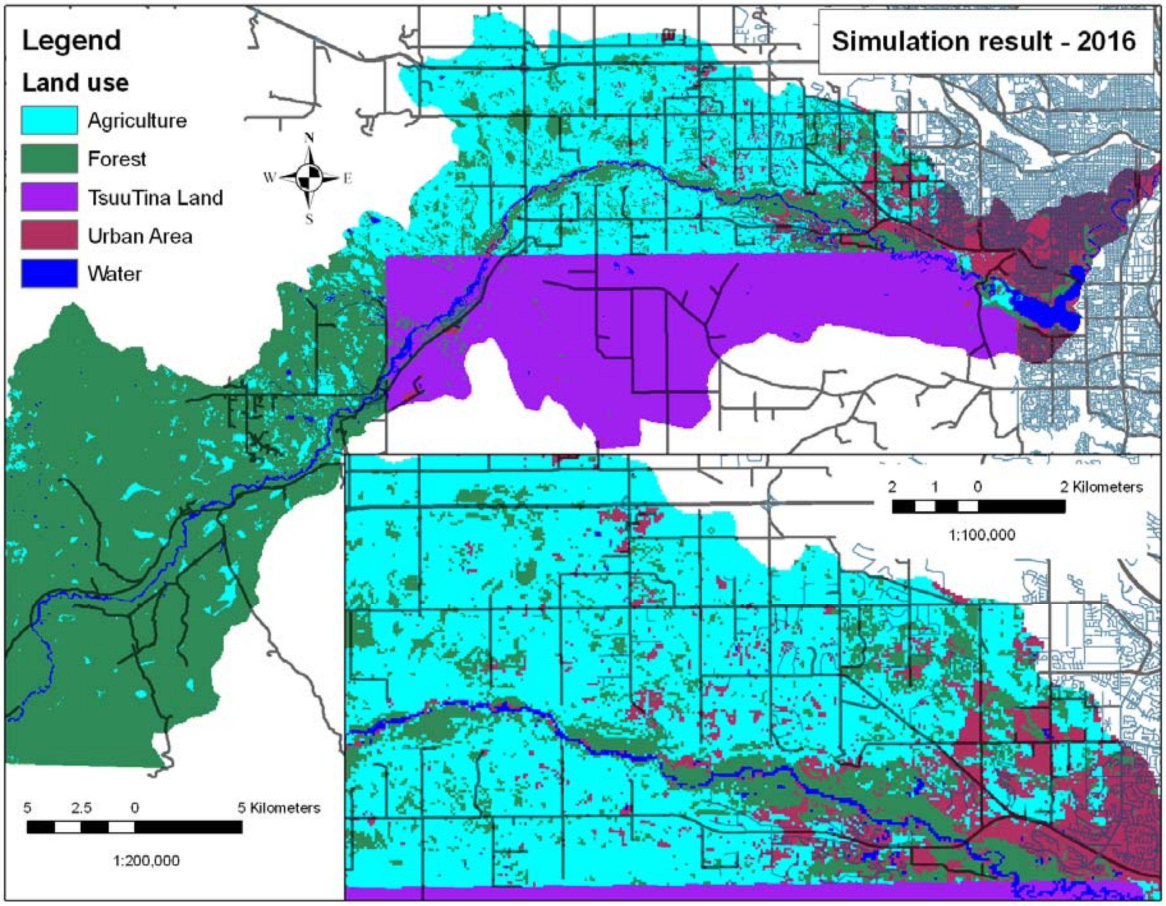

Figure 3.7. Simulation results for the year 2016. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006.

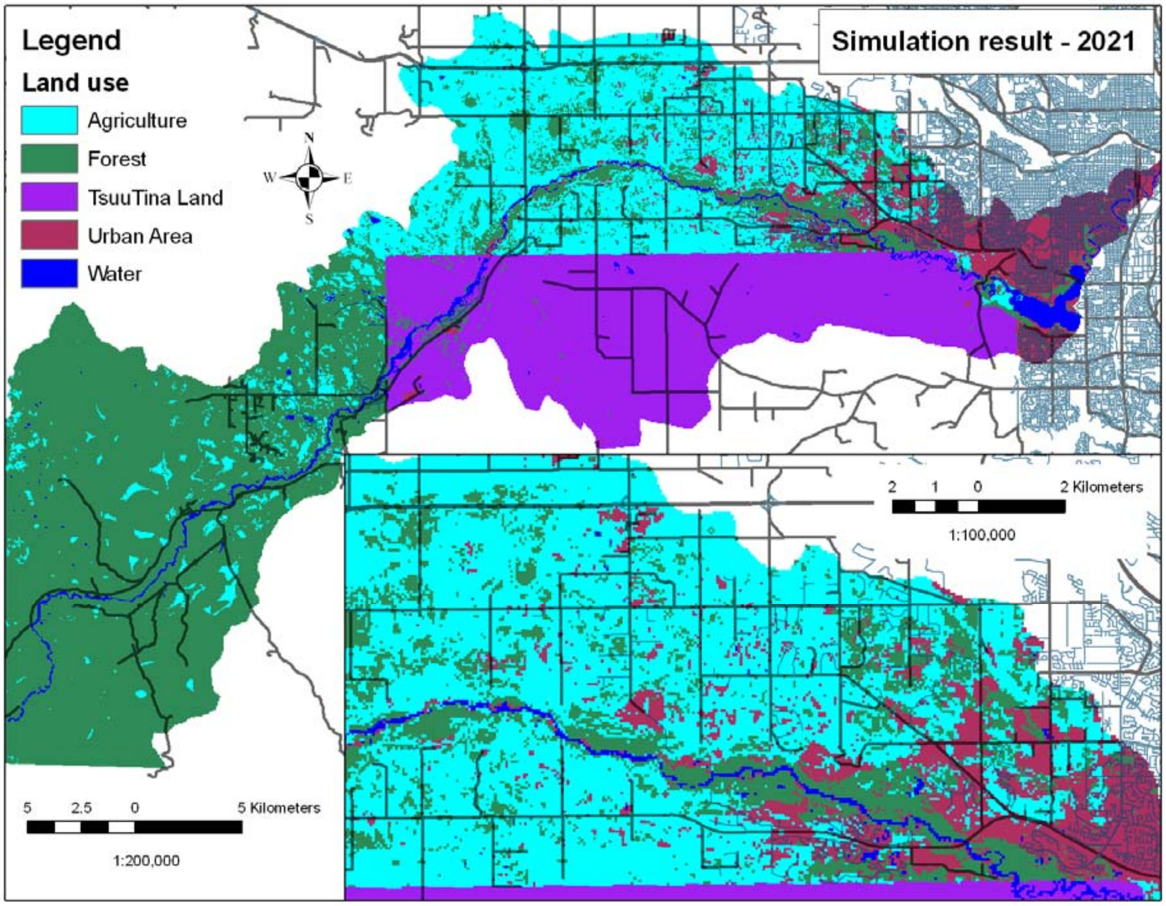

Figure 3.8. Simulation results for the year 2021. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006.

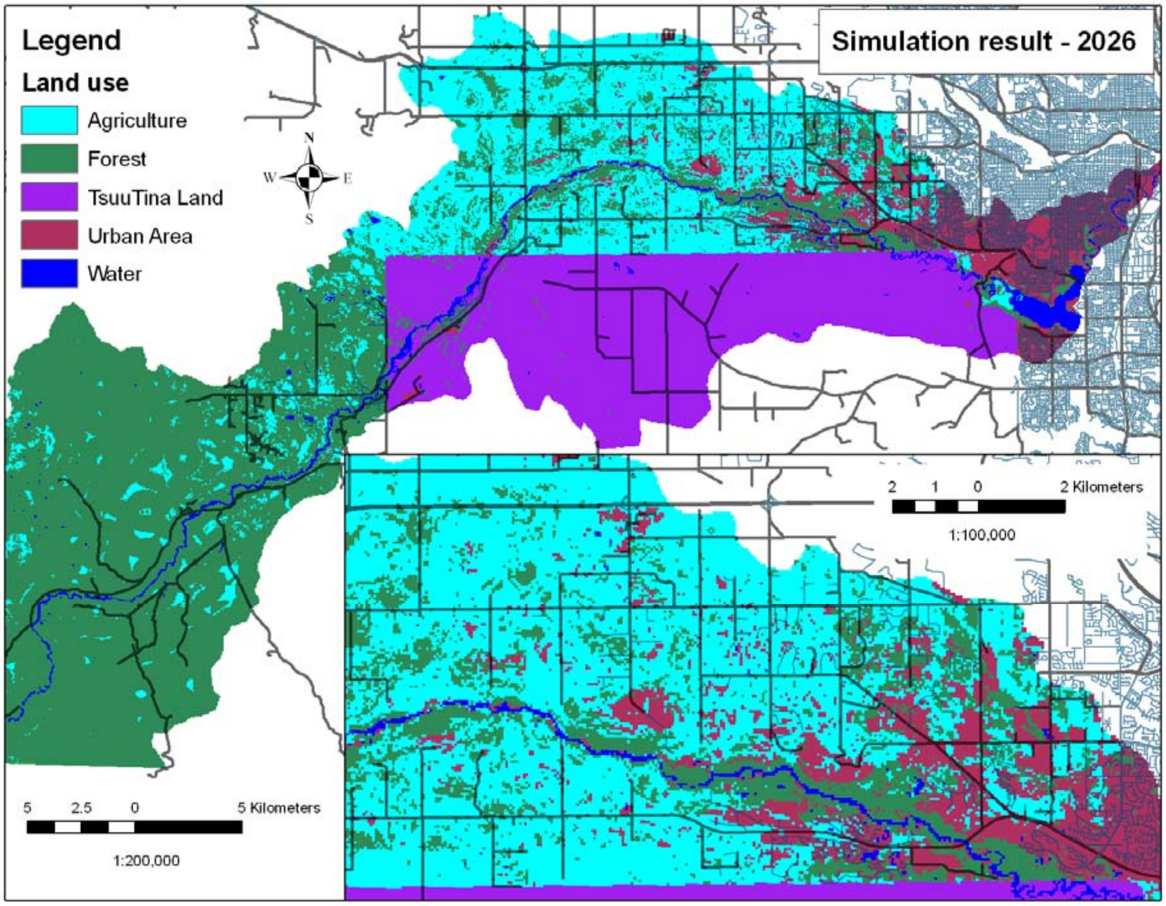

Figure 3.9. Simulation results for the year 2026. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006.

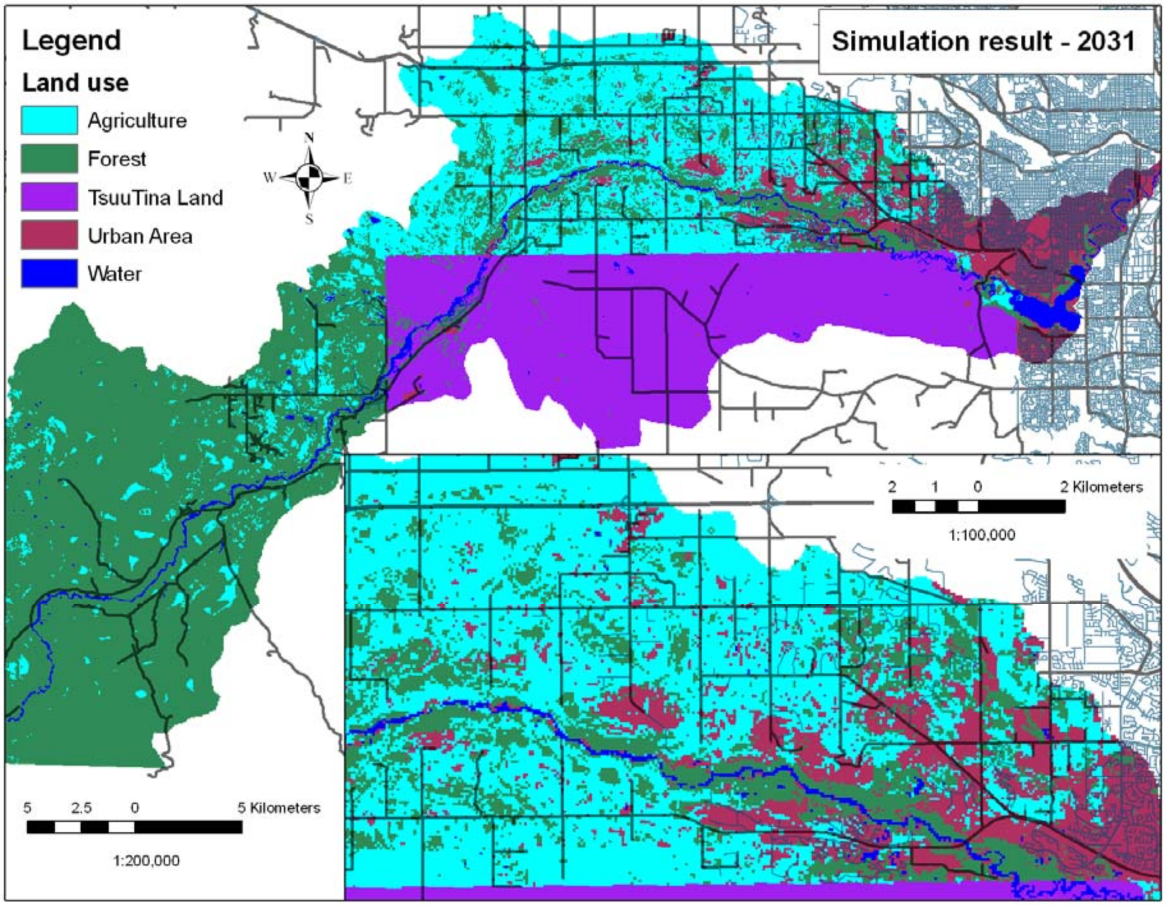

Figure 3.10: Simulation results for the year 2031. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006

Table 3.3: Percentage of the study area covered by different land uses in the simulation results from 2006 to 2031

	2011	2016	2021	2026	2031
Agriculture	24.93	24.46	24.00	23.53	23.07
Forest	46.37	46.33	46.28	46.23	46.18
Urban	5.32	5.83	6.35	6.86	7.38

According to the driving factors that were used, these results seem very plausible. The forest along the river remains, and tends to form small to medium sized patches elsewhere. Also, urban expansion mainly occurs close to Calgary and close to a main road. Urban development tends to appear in large patches, similar to what has been observed in the past.

The main urban developments along the Elbow River have been confirmed as actual urban development projects by experts from the validation group. Similarly, some much smaller developments represented by a few contiguous pixels also occur within the study area. The change from forest to agriculture is surprising on the western side of the study area as this area is nowadays mainly forested. It can be explained by the presence of such changes in the historical data sets. In a refined version of the model, data such as the type of soil could be used as an additional factor or constraint to limit the development of agriculture in this area. Experts from the validation group mentioned the presence of oat cultures in this area. The land use changes within the study area correspond well to the trend of changes identified in the historical land-use maps. Agriculture slowly disappears while forest tends to be stable and urban areas considerably increase.

To evaluate the influence of the choice and number of driving factors, a simulation was done without the driving factor “distance to a main river”. The simulation result obtained for 2031 (Figure 3.11) are very different from the results of the previous simulation (Figure 3.10). Urban developments are made by smaller patches and are scattered across the study area. Similarly, forest patches appear at different places and the sizes of the patches are much bigger. This means that distance to a main river is a key factor explaining land-use changes in the study area. In a broader perspective, it also means that the choice and number of driving factors is extremely important to simulate plausible land-use maps.

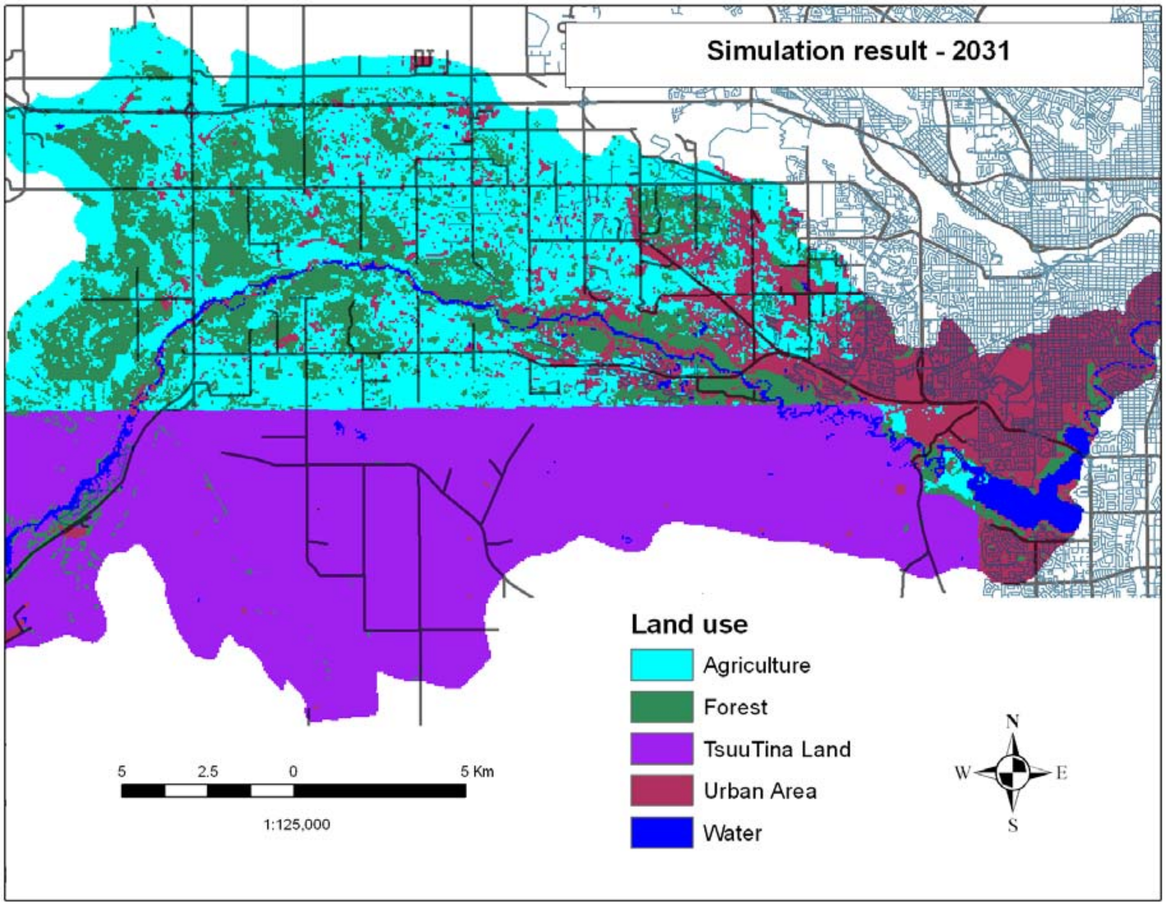

Figure 3.11. Simulation result for 2031. Only two driving factors have been considered, the distance to a main road and the distance to downtown Calgary

3.3.2 Scenario with population growth constraint

This section presents the results obtained when adding constraints to the model. The number of cells that should change state to urban state is specified to the model for each time step.

Since the correspondence between the outputs of the constrained simulation and the simulation described in the previous section is extremely high (between 97% and 99.98%), only the last result for 2031 is shown (Figure 3.12). This similarity between the results demonstrates that the method used in the model to establish the number of cells that should change state is valid and accurate.

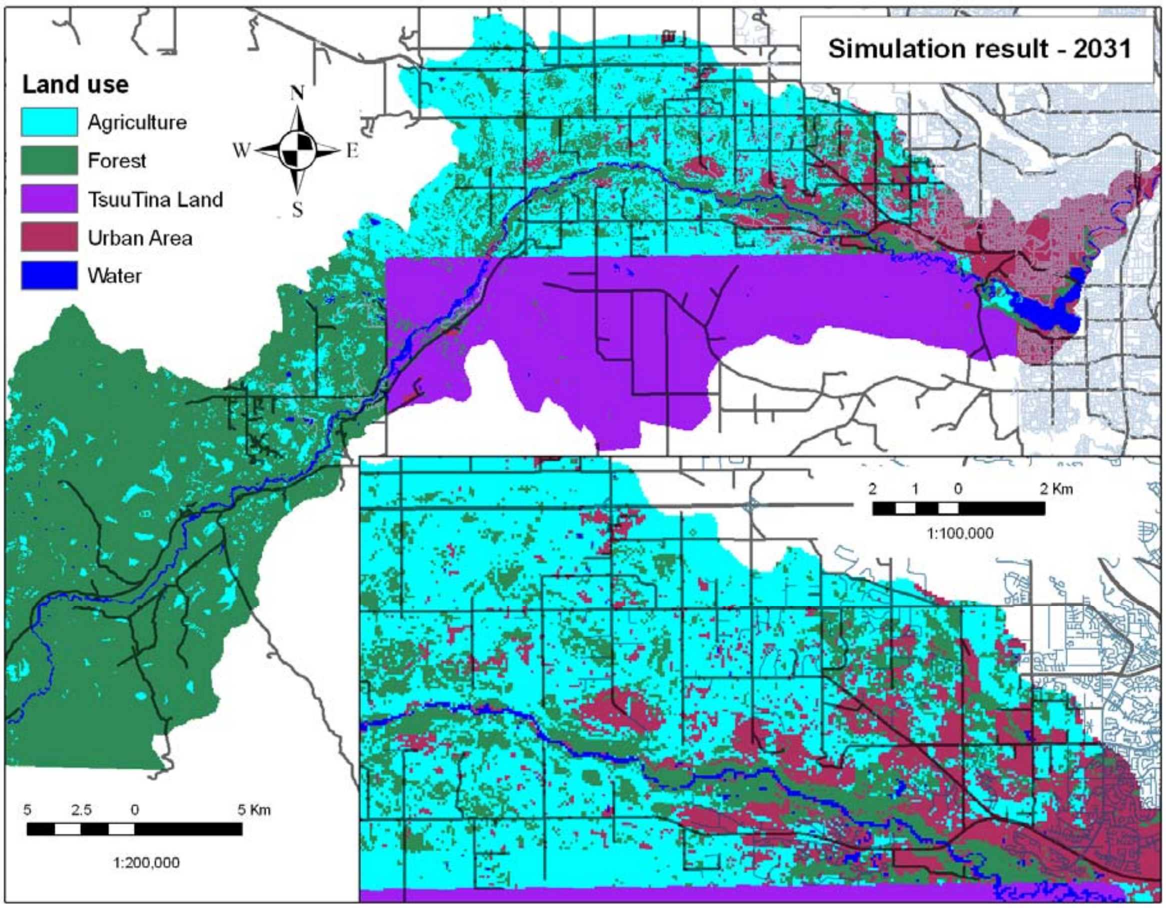

Figure 3.12. Simulation results for the year 2031. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

To test the effect of adding an extra land-use map to the calibration, the transition rules derived from the maps of 1985, 1992, 1996 and 2001 were used to run a simulation starting in 2001, for 2006, 2011, 2016, 2021, 2026 and 2031, constraining the number of cells changing their state to urban. The output for 2031 is shown on Figure 3.13. Most urban patches are located at the same place than when simulating with the transition rules derived with all the historical data, but the number of small urban patches is reduced. Moreover, the quantity and location of the forest patches are very different. This is due to the fact that different levels of intensity in land-use changes have been observed between the year 2001 and 2006. The model is still able to capture the important processes driving land-use change but can not know the new ones that appeared between 2001 and 2006. Consequently, during the calibration, the user should provide the model with data corresponding to years at which hitherto unseen processes were active over the study area or when processes were changing the land use at a very different intensity than the historical average. On the other hand, if the changes are caused by an exceptional event, like a major flood, it is better not to consider the corresponding land-use map as we do not want to simulate this exceptional event in the future, especially not at each simulated time step.

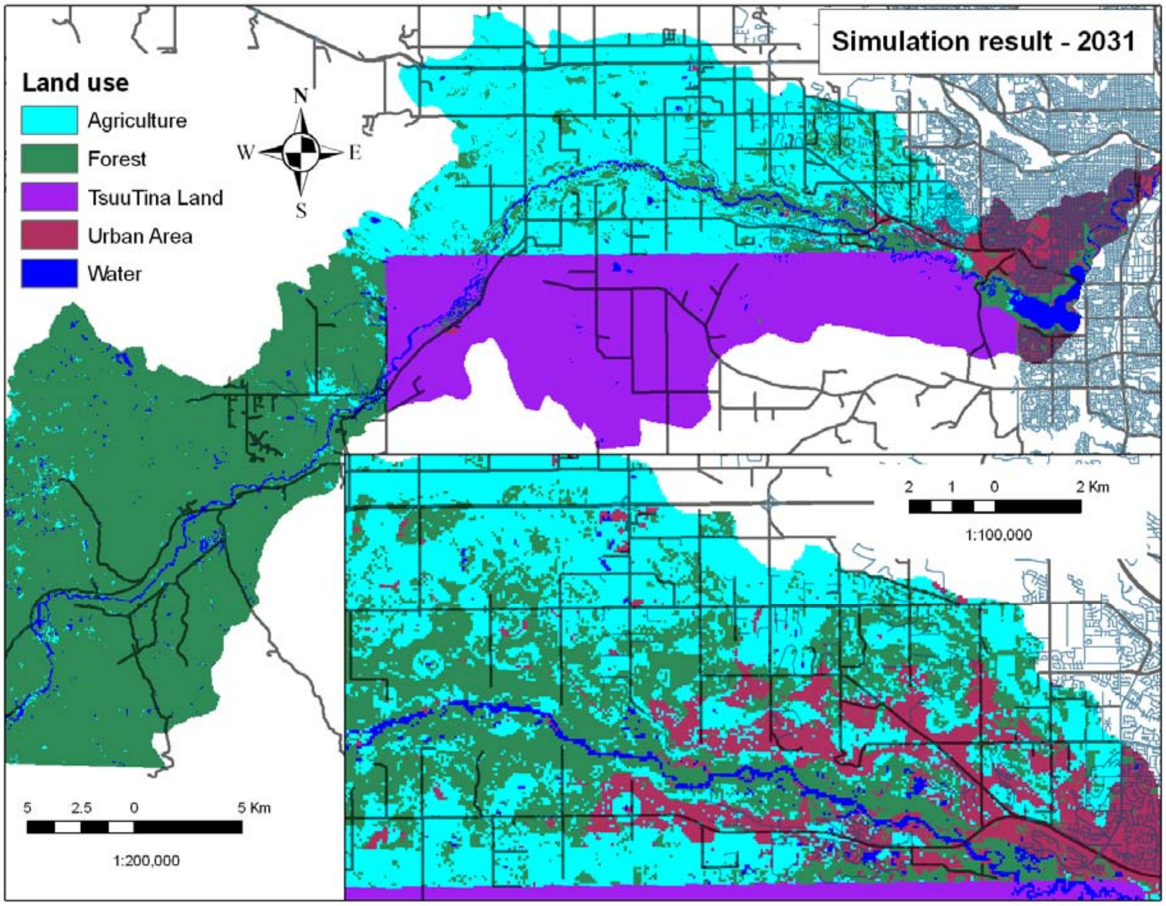

Figure 3.13. Simulation results for the year 2031. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996 and 2001. Urban growth is a constraint.

3.3.3 Scenario including a virtual new town

To demonstrate the ability of the CA model to simulate development scenarios, a modified 2006 land-use map was used as the initial conditions for the simulation. This land-use map displays a virtual new town in the northern part of the watershed, as shown on Figure 3.14.

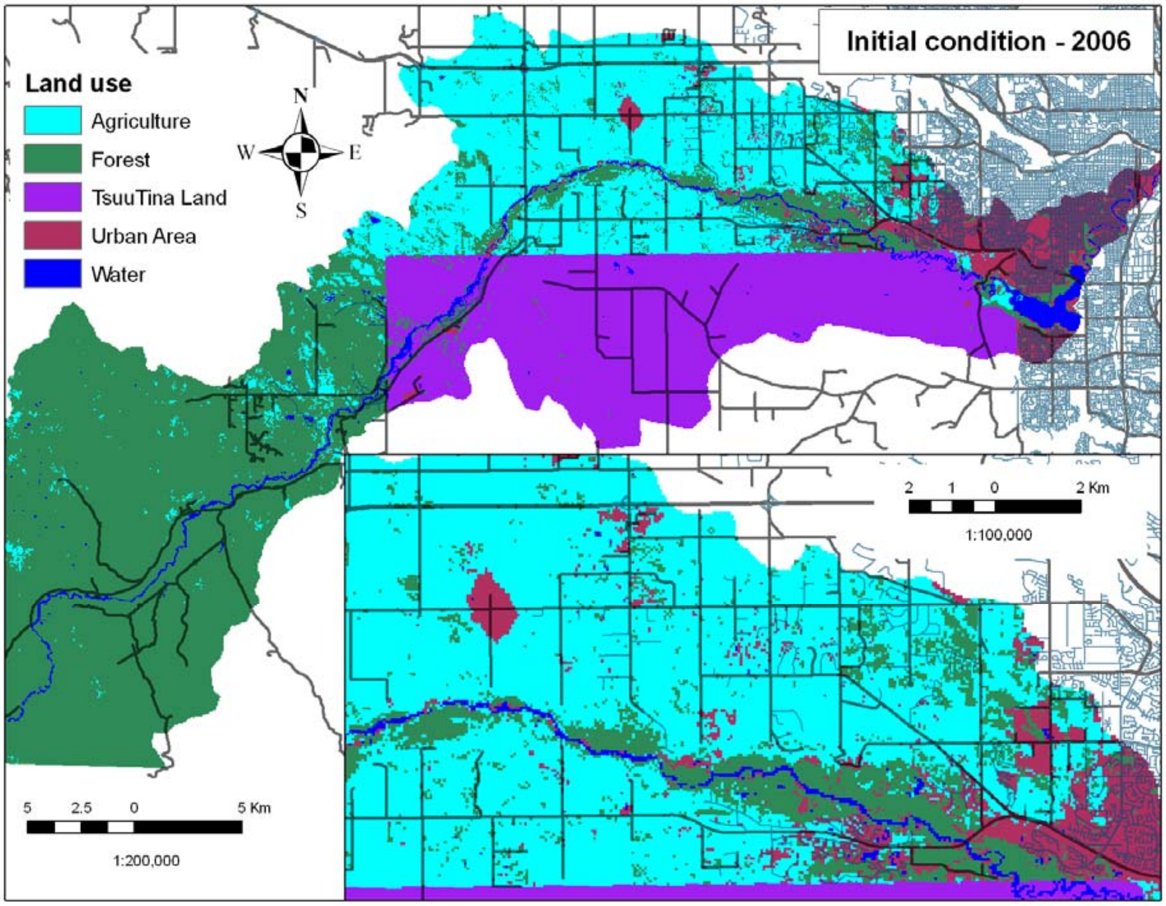

Figure 3.14. Modified 2006 land-use map displaying a virtual new town in the northern part of the watershed

In the first simulation, the modified 2006 map was used but the driving factor "distance to downtown Calgary" was kept as it is. The outputs of the simulation show little difference with the simulation ran with the non-modified initial condition map (Figure 3.15). This new town is located far away from Calgary and there is no transition rule describing the growth of a town far from the Calgary city limits. The impact of this virtual town is therefore minimal.

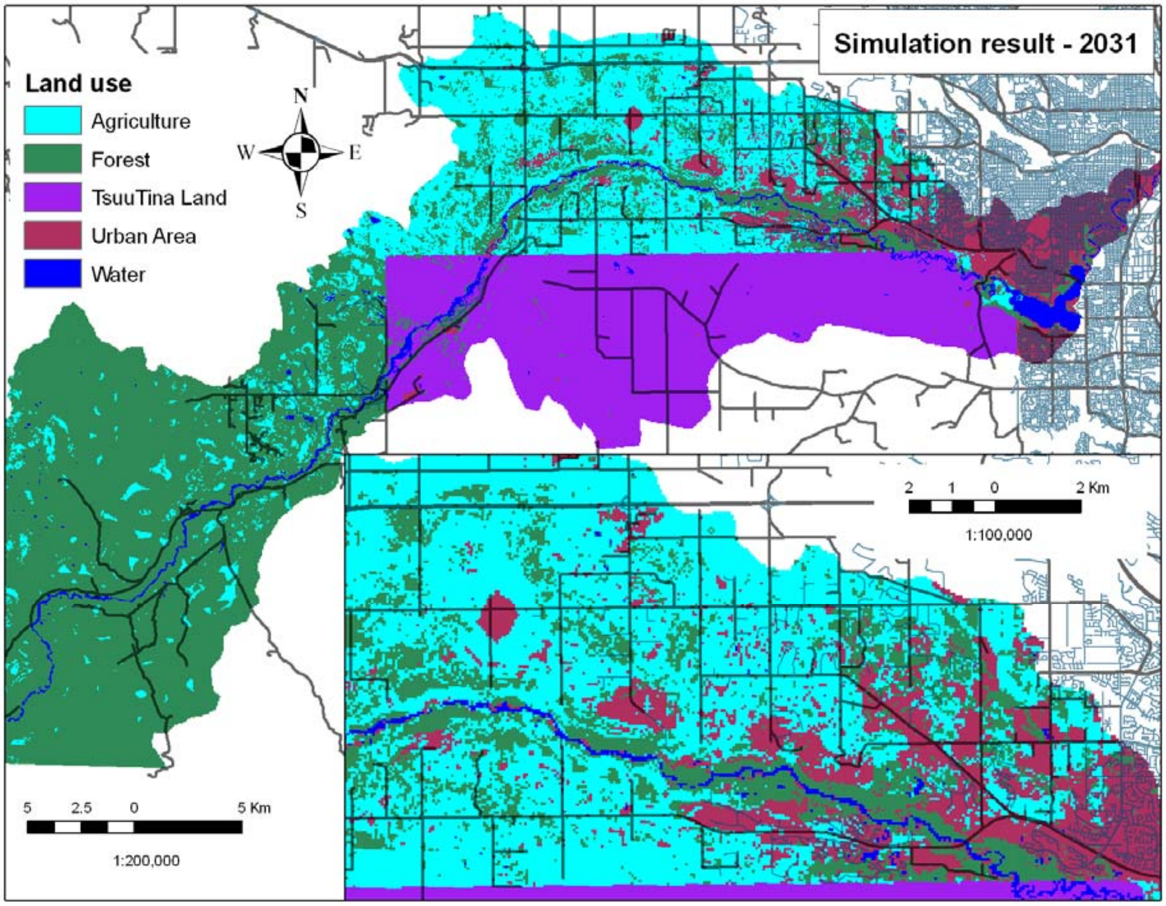

Figure 3.15. Simulation results for the year 2031 using the virtual new town in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

In the second simulation, the layer “distance to downtown Calgary” was replaced by the layer “distance to town centers”, which includes distances increasing from downtown Calgary and from the new town limits. Results are very different from the previous simulation and are represented for the years 2011, 2016, 2021, 2026 and 2031 on Figure 3.16, Figure 3.17, Figure 3.18, Figure 3.19 and Figure 3.20, respectively. The percentage of the study area covered by the different land uses is presented in Table 3.4.

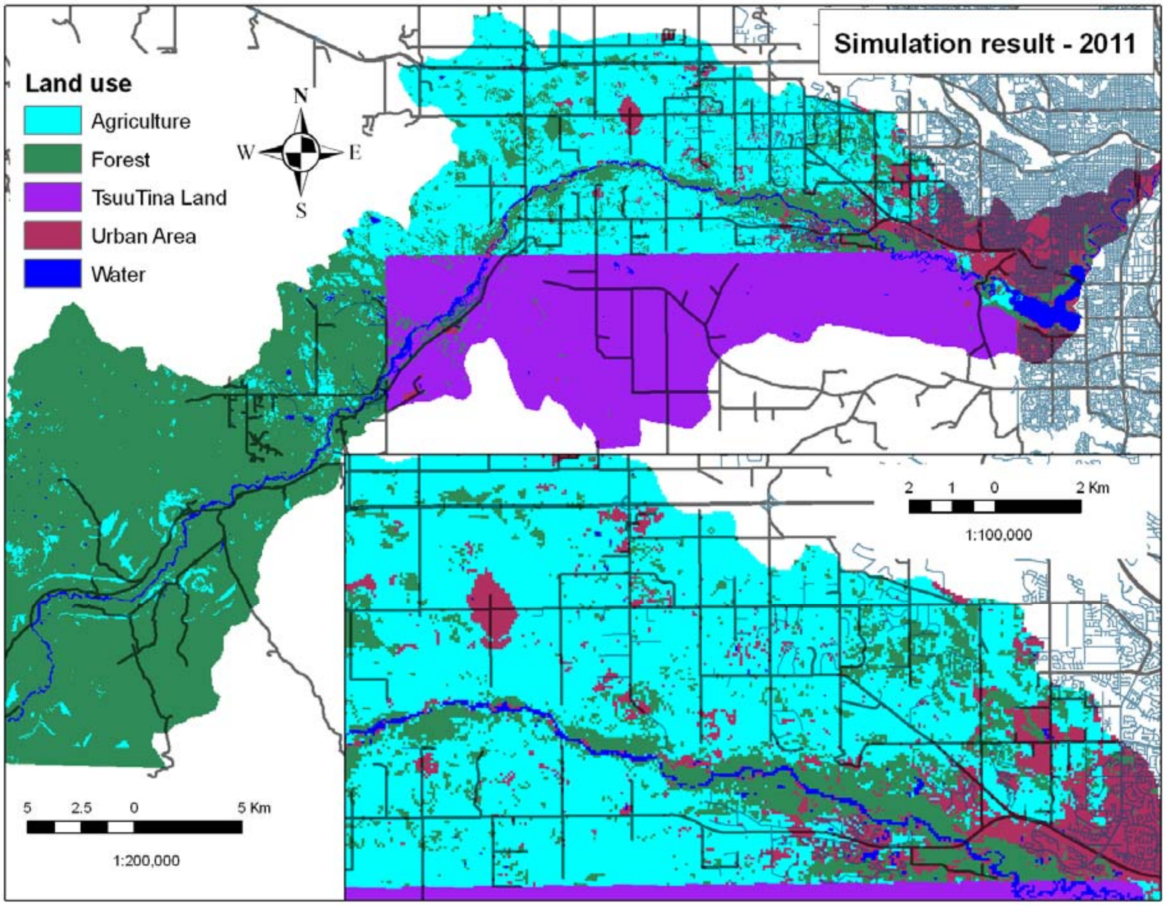

Figure 3.16. Simulation results for the year 2011 using the virtual new town and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

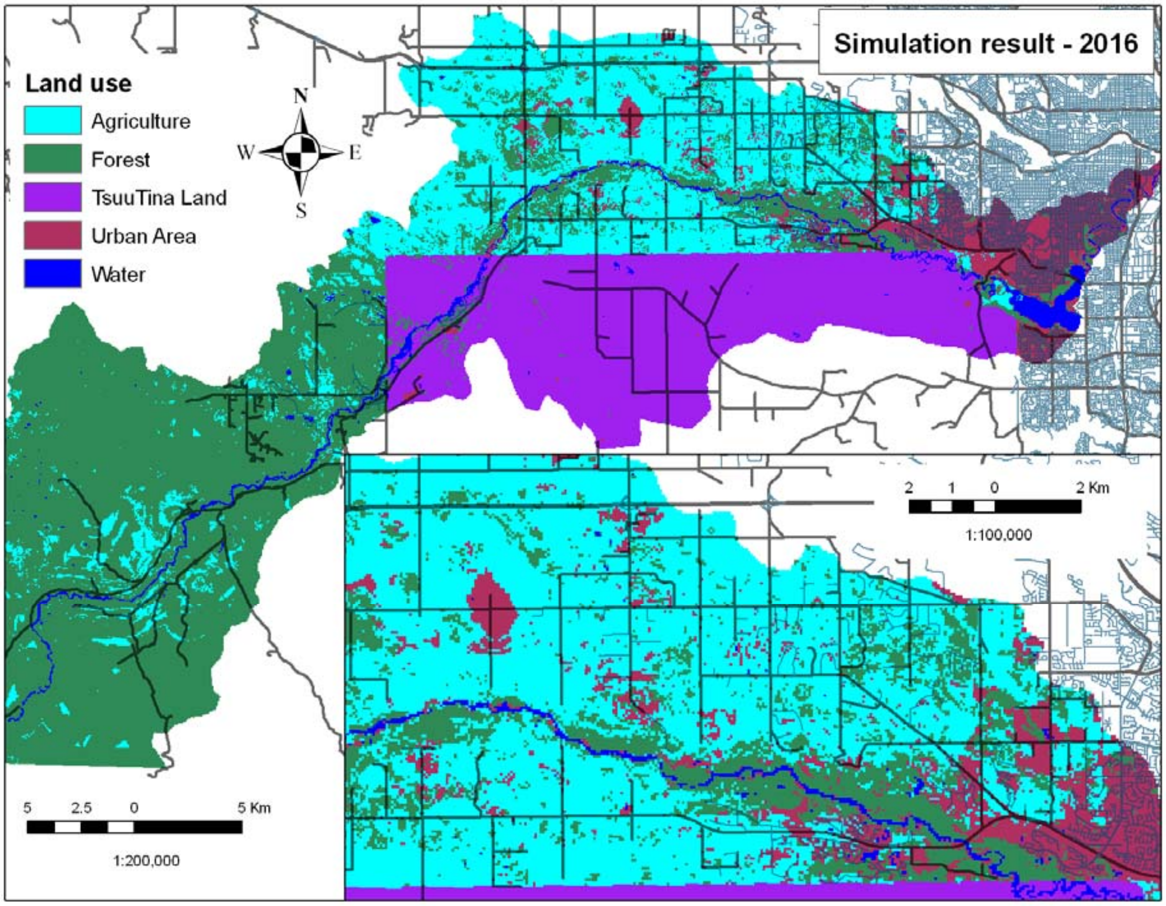

Figure 3.17. Simulation results for the year 2016 using the virtual new town and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

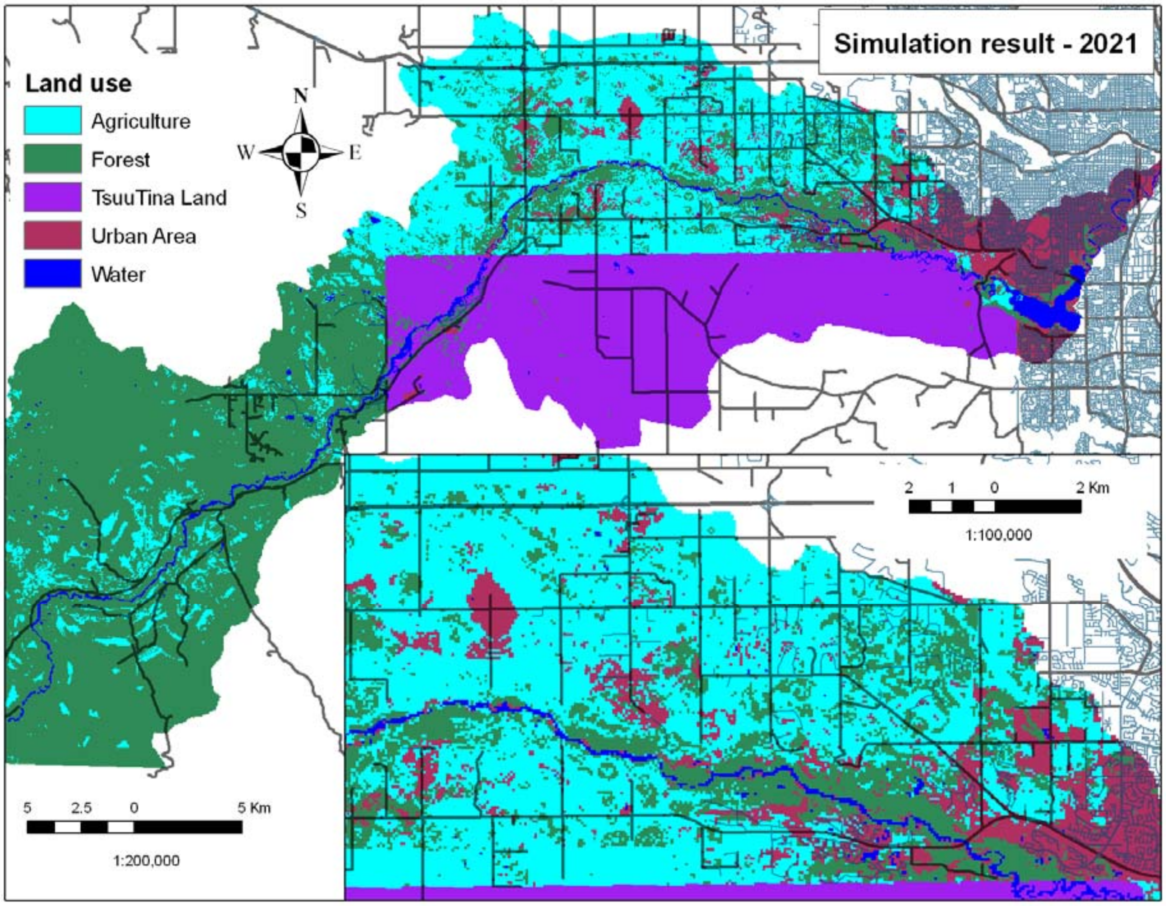

Figure 3.18. Simulation results for the year 2021 using the virtual new town and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

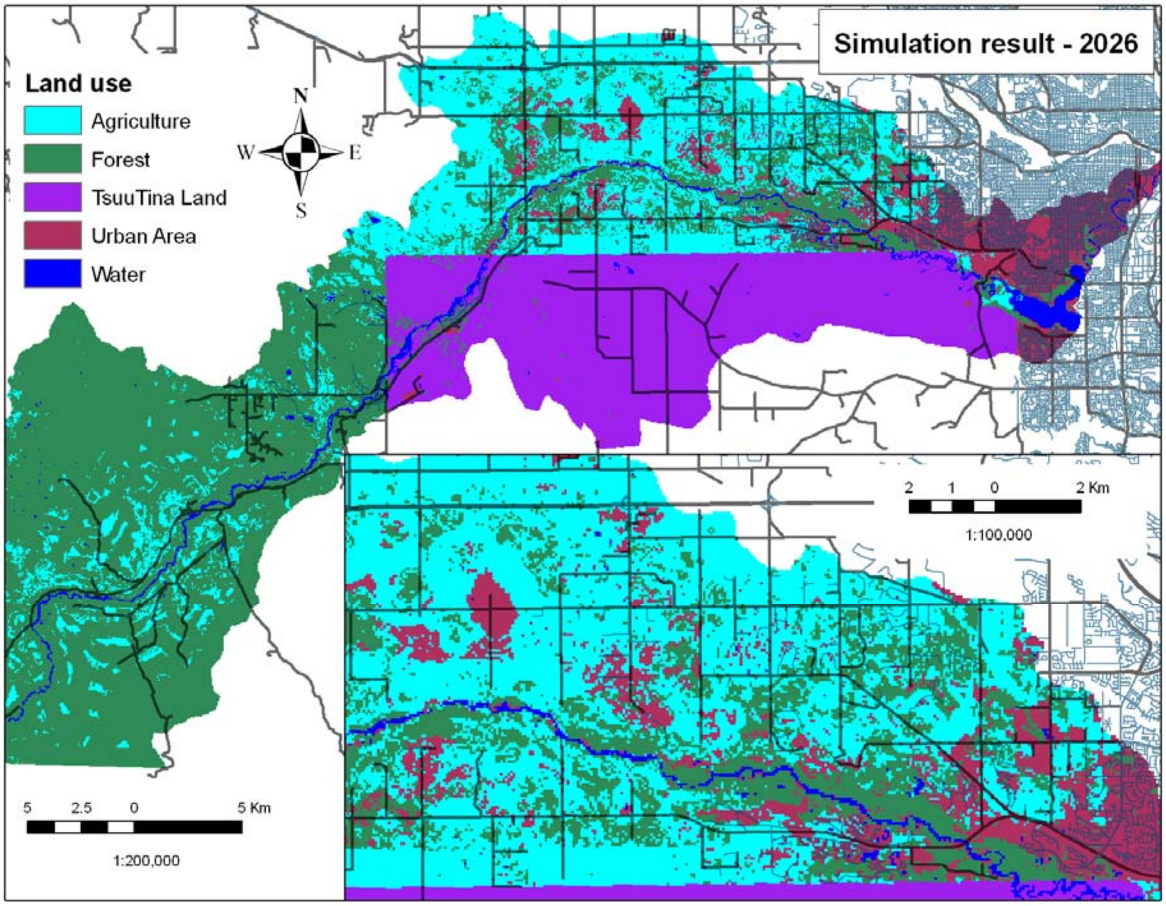

Figure 3.19. Simulation results for the year 2026 using the virtual new town and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

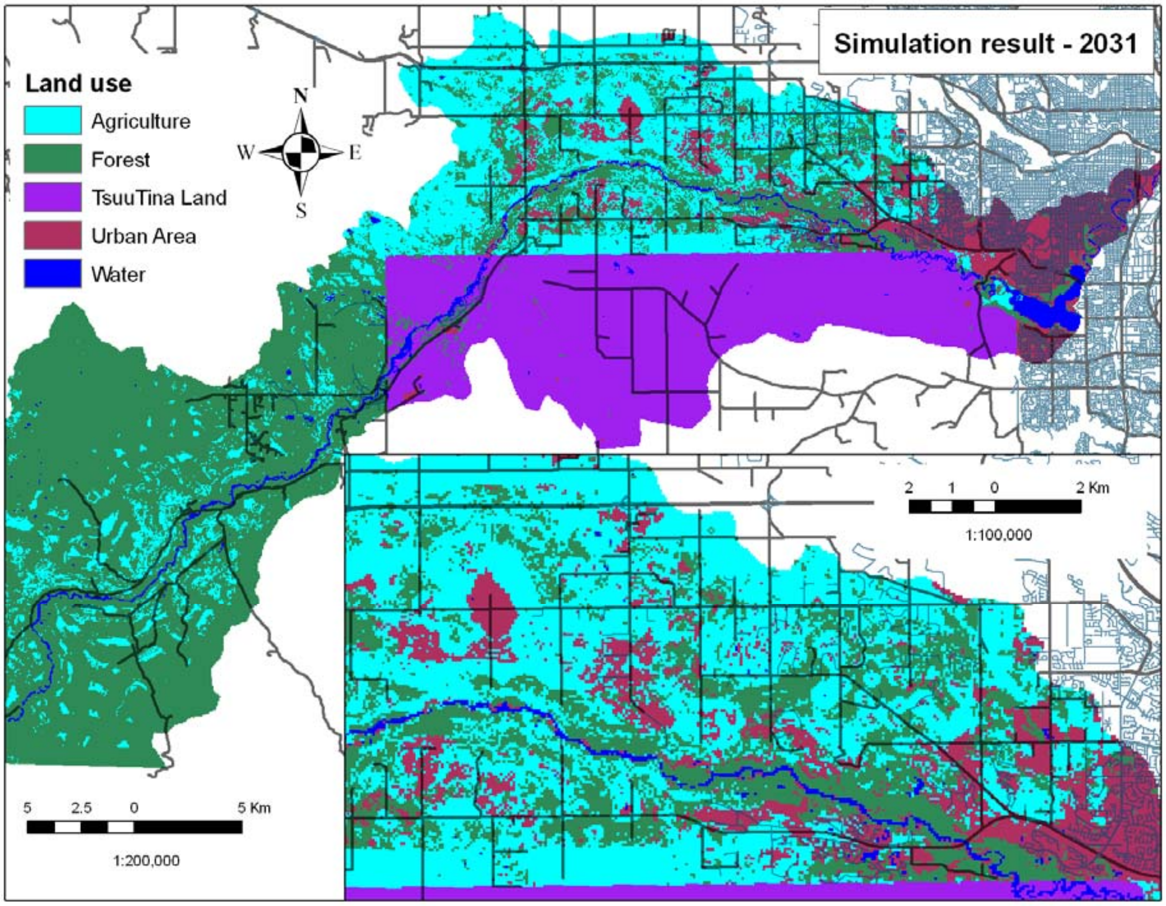

Figure 3.20. Simulation results for the year 2031 using the virtual new town and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

Table 3.4: Percentage of the study area covered by different land uses in the simulation results from 2006 to 2031 in presence of a new virtual town

	2011	2016	2021	2026	2031
Agriculture	24.72	24.25	23.78	23.32	22.85
Forest	46.37	46.32	46.28	46.23	46.18
Urban	5.53	6.05	6.56	7.07	7.59

It can be observed that the new town grows and that new urban developments are also being established in the vicinity of the new town. Developments between Bragg Creek and the new town, as well as between the Elbow River and the Tsuu Tina nation that were not present in the previous simulations, appear in these results. On the other hand, urban developments close to Calgary tend to slow down, which is expected as the total number of new urban cells is specified to the model at each time step. When comparing Tables 3.4 and 3.2, one can see that the urban land use always covers an additional 0.21% in presence of the new town. Also, the coverage of the forest, in terms of percentage of the study area, is the same in both tables. However, when looking at the spatial distribution of the forested cells in the simulation results that do and do not consider the new town, it appears that in 2031, 11 753 cells (7.5% of the study area) are not located at the same place. When the new town is considered, there are fewer forest cells in the western part of the watershed and, in the northern part of the watershed, forest cells appear in larger patches. One can interpret this result by saying that the new town becomes a regional attractive pole for residential development, which in turn affects the location of forest and agriculture in the whole watershed.

The last simulation was done using the original 2006 land-use map as initial conditions, therefore without the new town, but nonetheless using the driving factor layer "distance to town centers" instead of "distance to downtown Calgary". Results are similar to the ones just previously described so only the result corresponding to 2031 is shown (Figure 3.21). This similarity of results demonstrates the importance of the driving factors. The land uses in the neighborhood of a cell clearly have an important role in the land-use changes since there is no new urban development south of where the new town was established in the previous simulation. However, most of the new urban developments appear at the same location than in the previous simulation. One must therefore be careful to use different driving factor layers when these factors are directly influenced by the integration of a new feature in the landscape.

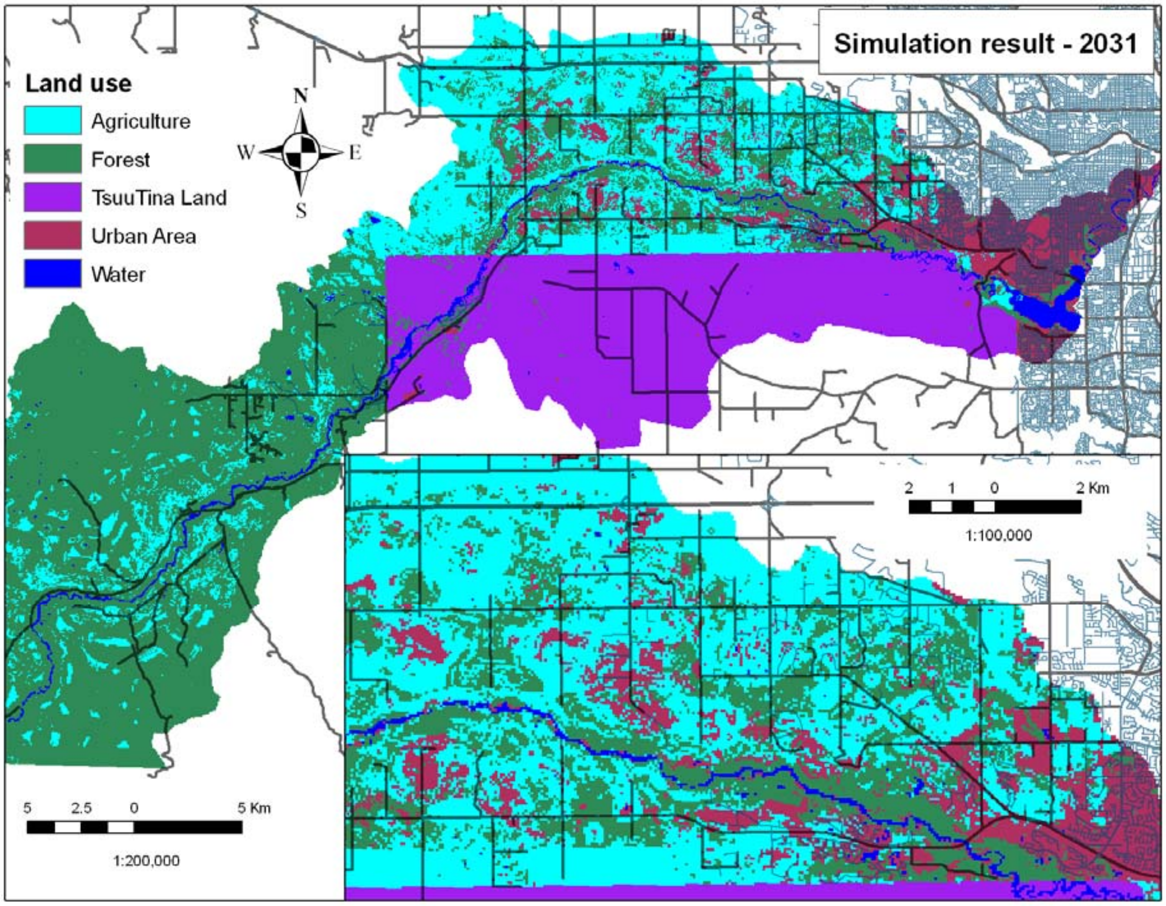

Figure 3.21. Simulation results for the year 2031 using the original 2006 land-use map and the distances to town centers in the initial condition. Resolution: 60 m; Land-use data used in the calibration: 1985, 1992, 1996, 2001 and 2006. Urban growth is a constraint.

Chapter 4: Conclusion

The objective of this research was to develop a new method to dynamically identify transition rules and calibrate a land-use CA model. This method had to be accurate, predictable, repeatable, adaptable and easily understandable by non-expert users. While calibrating the CA model, the user had to be able to learn about the processes driving land-use changes. Different types of constraints had to allow the user to simulate “what if” development scenarios. The developed CA model was then applied to simulate land-use changes in the Elbow River watershed over the next twenty-five years.

The results demonstrated that the objectives have been met in this thesis. The calibration procedure allows the transition rules to be dynamically created from the historical data, the neighborhood size and the selected driving factors. The variety of spatial conditions driving land-use changes were captured by thousands of transition rules. While most CA models use a limited number of rules to represent the dominant conditions driving the land-use changes, the CA model developed in this thesis considers several variations of these conditions. Simulations results are therefore very close to the reality, which is important for the planners using the model. By integrating constraints, the users can have a realistic idea of the evolution of the landscape. To perform a trustable calibration of the model, the historical land-use maps must be of good quality, must cover a period long enough and the date at which the data have been acquired must reflect the different spatial conditions that drive land-use changes. The developed calibration procedure allows a user to visually identify important neighborhood configurations that lead to a land-use change and to understand the importance of the driving factors as well as the presence of cells of other land uses within the neighborhood. The visual representation of these influences and the possibility to consider any driving factor, are novel features that help the user to understand the processes driving land-use changes. The transition rules are first described as conditional rules, which are easy to understand by the user, and are further internally transformed into mathematical rules, which are easily applied on historically unseen neighborhood configurations, in which case the transition rule associated with the most similar neighborhood configuration will be applied. This model is very adaptable as it can be applied over a study area of any shape, using data at any spatial and temporal resolution, considering any number of land uses and driving factors and can produce outputs at any wanted date. Moreover, the dynamic creation of the transition rules allows this model to be applied anywhere, making it a universal model. Application of the CA model from the past to the present demonstrated the validity and the accuracy of this approach, which provides better results than most other land-use CA models described in the literature. Simulations under constraints demonstrated the usefulness of this model for regional planners who wish to foresee the possible outcome of a development policy. The integration of a new virtual town led to urban development in locations where no developments have been observed before. Moreover, the attraction of this new town is different whether the distance to this new town is considered or not as a driving factor. By combining different constraints, the user can incorporate into the model a large number of potential development scenarios.

Of course, this CA model is in its first version and several improvements can be done to enhance the output quality. First, an automated way of finding the correct neighborhood ring sizes should be developed. One approach that is envisioned to do so is the use of 3-dimensional histograms having frequency on the Z axis, distance to central cell on the X axis and number of cells of the given land use at the given distance on the Y axis. It is expected that discrepancies will appear in this 3D "surface" that might represent the distances required for the rings. Secondly, it would be interesting to allow the development of patches of cells at once, rather than cell after cell in order to better mimic the appearance of new urban communities. Third, the influence of additional driving factors should be assessed. At last, the extrapolation method that is used to compute the number of cells that should adhere to each type of land-use change and to each rule should be changed for a more complex one that can find and perpetuate cyclic trends of changes. Improvement can also be done to the user interface.

To summarize, this CA model that represents about 22 500 lines of code, is a great tool that overcomes several limitations of current CA models, the main one being the ability to have multiple transition rules describing a land-use change. In this study, calibration was achieved in 45 minutes while simulating the land-use map at each time step required about 20 minutes. The application of constraints allows the user to simulate in a short amount of time complex scenarios, making this CA model a useful decision support tool. At last, this model can run on any desktop computer.

References

Abdelmoty, A. and B. El-Geresy (2000). Assessing spatial similarity in geographic databases. GIS and Geocomputation. P. Atkinson and D. Martin, Taylor and Francis**:** 255-269.

Alberta Finance Statistics. (2007, June 28, 2007). "Alberta Population Report, First Quarter, 2007." Retrieved July 9, 2007, from http://www.finance.gov.ab.ca/aboutalberta/population_reports/2007_1stquarter.pdf.

Batty, M. and Y. Xie (1994). "From cells to cities." Environment and Planning B: Planning and Design **21**: 31-48.

Batty, M. (2005). Cities and complexity : understanding cities with cellular automata, agent-based models, and fractals, MIT Press, 565 p.

Benenson, I. and P. M. Torrens (2004a). "Editorial: Geosimulation: object-based modeling of urban phenomena." Computers, Environment and Urban Systems **28**: 1-8.

Benenson, I. and P. M. Torrens (2004b). Geosimulation, Automata-based modeling of urban phenomena, John Wiley & sons,

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

Bettingera, P., M. Lennetteb, et al. (2004). "A hierarchical spatial framework for forest landscape planning." Ecological Modelling **182**: 25-48.

Beynon, M., S. Rasmequan, et al. (2002). "A new paradigm for computer-based decision support." Decision Support Systems **33**(2): 127-142.

Calgary Regional Partnership (2006). "Vision and mission statement."

Calgary Regional Partnership. (2007a). "Terms of Agreement for Working Together." Retrieved July 25, 2007, from http://www.calgaryregion.ca/crp/media/12854/terms%20of%20agreement%20-%20may%207%202007%20-%20with%20proposed%20ammendment.pdf.

Calgary Regional Partnership. (2007b). "Frequently Asked Questions." Retrieved July 25, 2007, from http://www.calgaryregion.ca/crp/aboutus/faqs.aspx.

Candau, J. (2002). Temporal calibration sensitivity of the SLEUTH urban growth model. Department of Geography, University of California. **Masters thesis**.

Caruso, G., M. Rounsevell, et al. (2005). "Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the Brussels periurban area." International Journal of Geographical Information Science **19**(2): 103-123.

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

Cheng, J. and I. Masser (2004). "Understanding spatial and temporal processes of urban growth: cellular automata modelling." Environment and Planning B: Planning and Design **31**: 167-194.

City of Calgary (2006). Socio-Economic Outlook 2006-2016. Calgary.

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

Department of Planning and Community Services M.D. of Rocky View. (2007, January 2007). "Harmony - Conceptual Scheme." Retrieved July 9th, 2007, from http://www.rockyview.ca/uploaded/harmonycsfirstreadingdraft.pdf.

Dietzel, C. and K. C. Clarke (2006a). "The effect of disaggregating land use categories in cellular automata during model calibration and forecasting " Computers, Environment and Urban Systems **30**(1): 78-101.

Dietzel, C. and K. C. Clarke (2006b). "Decreasing Computational Time of Urban Cellular Automata Through Model Portability " GeoInformatica **10**(2): 197-211.

Dietzel, C. and K. C. Clarke (2007). "Toward Optimal Calibration of the SLEUTH Land Use Change Model." Transactions in GIS **11**(1): 29-45.

ESRI (2005). "resampling, methods." Desktop help.

Fang, S., G. Z. Gertner, et al. (2005). "The impact of interactions in spatial simulation of the dynamics of urban sprawl." Landscape and Urban Planning **73**: 294-306.

Gardner, M. (1970). "The fantastic combinations of John Conway's new solitaire game "life"." Scientific American **223**: 120-123.

Georgoudas, I. G., G. C. Sirakoulis, et al. (2007). "Modelling earthquake activity features using cellular automata." Mathematical and Computer Modelling **46**(1-2): 124-137.

Gov. of Alberta. (2007, June 18, 2007). "Land-use Framework." Retrieved July 25, 2007 2007, from http://www.landuse.gov.ab.ca/.

Gutowitz, H. A. (1990). "Introduction." Physica D **45**: vii-xiv.

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

He, C., N. Okada, et al. (2006). "Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China." Applied Geography **26**(3-4): 323-345.

ITTVIS (2007). IDL.

Jantz, C. A., S. J. Goetz, et al. (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: Planning and Design **30**: 251-271.

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

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

Kocabas, V. and S. Dragicevic (2004). Sensitivity analysis of a GIS-based cellular automata model. Int. archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ed. By Altman M.O, XXth ISPRS Congress, Istanbul, Turkey, 12-23 July 2004, Vol XXXV, part B.

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

Lett, C., C. Silber, et al. (1999). "Comparison of a cellular automata network and an individual-based model for the simulation of forest dynamics." Ecological Modelling **121**: 277-293.

Li, X. and A. G.-O. Yeh (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. and A. G.-O. Yeh (2002). "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. and A. G.-O. Yeh (2004). "Data mining of cellular automata's transition rules." International Journal of Geographical Information Science **18**(8): 723-744.

Liu, X. and C. Andersson (2004). "Assessing the impact of temporal dynamics on land-use change modeling." Computers, Environment and Urban Systems **27**: 107-124.

Liu, Y. and s. R. Phinn (2001). Developing a cellular automaton model of urban growth incorporating fuzzy set approaches. 6th int. conf. on Geocomputation, U of Queensland, Brisbane, Australia 24-26 sept 2001.

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

Ménard, A. and D. J. Marceau (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**(3-4): 253-265.

Messina, J. P., S. J. Walsh, et al. (1999). The application of cellular automata modeling for enhanced landcover classification in the Ecuadorian Amazon. Proc. of the 4th Int conf. on GeoComputation, Mary Washington College, Fredericksburg, VA, USA 25-28 July 1999.

Moreno, N. and D. J. Marceau (2006). A Vector-Based Geographical Cellular Automata Model To Allow Polygon Shape Change. Summer Simulation Multiconference Calgary, Ab.

Municipal Affairs and Housing; Gov. of Alberta. (2007, July 16, 2007). "Government of Alberta response to report of the Minister's Council on Municipal Sustainability." Retrieved July 24, 2007, from http://www.gov.ab.ca/acn/200707/21792CF9902ED-08E9-D748-402F6CEDD23CEC6C.html.

O'Sullivan, D. (2001). "Exploring spatial process dynamics using irregular cellular automaton models." Geographical Analysis **33**(1).

Ohgai, A., Y. Gohnai, et al. (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.

Parrott, L. and R. Kok (2000). "Incorporating Complexity in Ecosystem Modelling." Complexity international **7**.

Parwani, R. R. (2002). "Complexity: an introduction." Complexity international **Draft Manuscript**: 1-9.

Pettit, C. J. (2005). "Use of a collaborative GIS-based planning-support system to assist in formulating a sustainable-development scenario for Hervey Bay, Australia." Environment and Planning B: Planning and Design **32**(4): 523-545.

Railsback, S. F. (2001). "Concepts from complex adaptive systems as a framework for individual-based modelling." Ecological Modelling **139**: 47-62.

Remington, R. (2006). Town of 5,000 Proposed. Calgary Herald.

Richards, J. A. (2006). Remote sensing digital image analysis : an introduction Springer, 439

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.

Shi, W. and M. Y. C. Pang (2000). "Development of Voronoi-based cellular automata—an integrated dynamic model for Geographical Information Systems." International Journal of Geographical Information Science **14**(5): 455-474.

Soares-Filho, B. S., G. C. Cerqueira, et al. (2002). "DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier." Ecological Modelling **154**: 217-235.

Statistics Canada. (2007a). "The Daily - Canada's population estimates." Retrieved September 29, 2007, from http://www.statcan.ca/Daily/English/070927/d070927a.htm.

Statistics Canada (2007b). 2006 Canadian Census.

Stevens, D., S. Dragicevic, et al. (2007). "iCity: A GIS–CA modelling tool for urban planning and decision making " Environmental Modelling & Software **22**(6): 761-773.

Straatman, B., R. White, et al. (2004). "Towards an automatic calibration procedure for constrained cellular automata." Computers, Environment and Urban Systems **28**: 149-170.

Sui, D. Z. and H. Zeng (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.

Takeyama, M. and H. Couclelis (1997). "Map dynamics: integrating cellular automata and GIS through Geo-Algebra." International Journal of Geographical Information Science **11**(1): 73-91.

Torrens, P. M. and D. O'Sullivan (2000). Cities, cells, and complexity: developing a research agenda for urban geocomputation. Proceedings of the 5th International Conference on GeoComputation, University of Greenwich, United Kingdom, 23 - 25 August 2000.

Torrens, P. M. and I. Benenson (2005). "Geographic Automata Systems." International Journal of Geographical Information Science **19**(4): 385-412.

Verburg, P. H., T. C. M. de Nijs, et al. (2004). "A method to analyse neighbourhood characteristics of land use patterns." Computers, Environment and Urban Systems **28**(6): 667-690.

Ward, D. P., A. T. Murray, et al. (1999). An optimized cellular automata approach for sustainable urban development in rapidly urbanizing regions. Proc. of the IV International Conference on GeoComputation, Mary Washington College, Fredericksburg, VA, USA, on 25-28 July 1999.

Ward, D. P., A. T. Murray, et al. (2000). "A stochastically constrained cellular model of urban growth." Computers, Environment and Urban Systems **24**(6): 539-558.

White, R. and G. Engelen (1993). "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns." Environment and Planning A **25**: 1175-1199.

White, R., G. Engelen, et al. (1997). "The use of constrained Cellular Automata for high-resolution modelling of urban land-use dynamics." Environment and Planning B: Planning and Design **25**: 323-343.

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

White, R. (2006). "Pattern based map comparisons." Journal of Geographical Systems **8**(2): 145-164.

Wolfram, S. (1983). "Statistical Mechanics of Cellular Automata." Reviews of Modern Physics **55**(3): 601-644.

Wolfram, S. (1984). "Preface." Physica D **10**: vii-xii.

Wolfram, S. (1988). Complex system theory. Emerging Syntheses in Science: Proceedings of the Founding Workshops of the Santa Fe Institute, (Addison-Wesley, 1988) 183-189.

Wu, F. (1998). "An experiment on the generic polycentricity of urban growth in a cellular automatic city." Environment and Planning B: Planning and Design **24**: 323-343.

Wu, F. (2000). A parameterised urban cellular model combining spontaneous and self-organising growth. GIS and Geocomputation. P. Atkinson and D. Martin, Taylor and Francis**:** 73-85.

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.

Yang, X. and C. P. Lo (2003). "Modelling urban growth and landscape changes in the Atlanta metropolitan area." International Journal of Geographical Information Science **5**: 463-488.

Yeh, A. G.-O. and X. Li (2001). "A constrained CA model for the simulation and planning of sustainable urban forms by using GIS." Environment and Planning B: Planning and Design **28**: 733-753.

Yeh, A. G.-O. and X. Li (2005). "Errors and uncertainties in urban cellular automata." Computers, Environment and Urban Systems **IN PRESS**.