



UCGE Reports
Number 20365

Department of Geomatics Engineering

A Cellular Automata Model to Simulate Land-use Changes
at Fine Spatial Resolution

(URL: <http://www.geomatics.ucalgary.ca/graduatetheses>)

by

Fang Wang

Dec. 12, 2012



UNIVERSITY OF CALGARY

A Cellular Automata Model to Simulate Land-use Changes at Fine Spatial Resolution

by

Fang Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

DECEMBER, 2012

© Fang Wang 2012

Abstract

Cellular Automata have become an important tool for modeling urban growth and land-use changes in recent decades. However, few studies have evaluated the possibility of building a CA model at fine spatial resolution for understanding detailed land-use dynamics. This research was carried out to develop a fine-resolution CA model, with an emphasis on tackling two main challenges. The first challenge is that the model complexity dramatically increases due to the increased number of land-use classes and driving factors at fine resolutions. In this thesis, Rough Set Theory was applied to guide the selection of dominant driving factors for the model calibration, resulting in a similar or improved performance as compared to the use of all factors. The second challenge involves designing a suitable CA model for land-use modeling at fine spatial resolution. Traditional cell-based CA are unable to generate reliable results at such resolutions because single cells often only represent components of land-use entities (i.e. houses or parks in urban residential areas), while recently proposed entity-based CA models usually ignore the internal heterogeneity of the entities. This thesis proposes a novel patch-based CA model to simulate land-use changes where the real-world entities are represented as patches. A patch refers to a collection of adjacent cells that, when combined together, represent an entity differing from its surroundings in nature or appearance. The results reveal that the patch-based CA model generates compact and realistic land-use patterns as found in the historical land-use maps. Calibrated in the eastern Elbow River watershed (adjacent to the City of Calgary), this model was further applied to simulate future land use in this area under three different development scenarios: *business-as-usual scenario*, *protective growth scenario* and *smart growth scenario*. The results reveal that both the *protective growth scenario* and the *smart growth scenario* consume less non-developed lands (i.e. agriculture and forest) than the *business-*

as-usual scenario does. The resulting maps generated by the patch-based CA model clearly illustrate that increasing land-use efficiency is an effective way to reduce the impact caused by a rapid population growth.

Acknowledgements

I would like to take this opportunity to express my deep and sincere gratitude to my supervisor Dr. Danielle Marceau. Her wisdom, knowledge and commitment to the highest standards inspired and motivated me. I would also like to thank my supervisory committee members Dr. Xin Wang and Dr. Andrew Hunter for their valuable advice and assistance in my research at various stages.

Many thanks to all my colleagues at the Geocomputing laboratory for their intellectual support and feedback regarding this thesis. Special thanks are due to Nishad Wijesekara, Babak Farjad, Majeed Pooyandeh, Dr. Niandry Moreno, and Jean-Gabriel Hasbani for their valuable discussions and suggestions about this research.

I am grateful to several agencies that funded this project: Alberta Innovates Technology Futures, University of Calgary, Calgary Regional Partnership, and NSERC (Natural Sciences and Engineering Research Council).

To my wonderful husband, this work cannot be completed without your unconditional love. You have supported me in the darkest time and believed in me even when I did not believe in myself. To my daughter - Claire, I am very sorry for the time I missed while you are growing. You are the best thing that I have ever had in my life. I love you. And thanks to my family in China for all their love and support.

Dedication

To Claire, Gang and my family in China.

Table of Contents

Abstract	ii
Acknowledgements	iv
Dedication	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
CHAPTER ONE: INTRODUCTION	1
1.1 Models for studying land-use change	1
1.2 Cellular Automata for land-use change modelling	6
1.2.1 Basic concepts of CA	6
1.2.2 CA-based land-use models	9
1.3 The rationale for studying land-use changes at fine spatial resolutions	11
1.4 Challenges related to developing a land-use CA model at fine spatial resolution ..	13
1.4.1 Factor selection in CA modeling	13
1.4.2 Space representation in CA models	16
1.5 Objective of the thesis	20
1.6 Assumptions and limitations	21
1.7 Organization of the thesis	21
CHAPTER TWO: IDENTIFYING DOMINANT DRIVING FACTORS USING ROUGH SET THEORY	24
2.1 Study area and datasets	24
2.2 Fundamentals of Rough Set Theory for feature selection	26
2.3 Procedure for identifying the dominant driving factors using RST	29
2.3.1 Decision table preparation	30
2.3.2 Selection of the driving factors using RST	36
2.3.3 Generating and implementing the CA model transition rules	38
2.3.4 Model simulation	39
2.3.5 Results comparison	41
CHAPTER THREE: DEVELOPMENT OF A PATCH-BASED CA TO SIMULATE LAND- USE CHANGES AT FINE RESOLUTION	45
3.1 Study area and dataset	45
3.1.1 Study area	45
3.1.2 Data sources	46
3.1.3 Land-use map generation	47
3.1.3.1 Image fusion and georegistration	47
3.1.3.2 Image segmentation and classification	48
3.2 Framework of the patch-based CA model	51
3.2.1 Calibration of the patch-based CA model	53
3.2.1.1 Identification of driving factors	53
3.2.1.2 Calibration using WOE	56
3.2.2 Simulation module of the patch-based CA model	60

3.2.2.1 Transition probability map combination.....	60
3.2.2.2 Patch-based simulation	61
3.2.3 Development of a cell-based CA model.....	62
3.2.4 Evaluation of the results	62
CHAPTER FOUR: SCENARIO EVALUATION WITH THE PATCH-BASED CA MODEL	
.....	64
4.1 Context of study.....	64
4.2 Data.....	69
4.3 Scenarios definition	71
4.3.1 Business-as-usual scenario	72
4.3.2 Protective growth scenario	73
4.3.3 Smart growth scenario.....	75
4.4 Simulation.....	77
CHAPTER FIVE: RESULTS AND INTERPRETATION	78
5.1 Factor selection using RST	78
5.1.1 Results	78
5.1.2 Summary.....	82
5.2 Development of the patch-based CA model	83
5.2.1 Results	83
5.2.2 Summary.....	92
5.3 Scenario Evaluation with the Patch-based CA Model.....	93
5.3.1 Projected area for each land-use class	94
5.3.2 Visual comparison	98
5.3.3 Land consumption	102
5.3.4 Summary.....	104
CHAPTER SIX: CONCLUSION AND FUTURE WORK.....	106
6.1 Thesis contributions.....	109
6.2 Future work.....	110
REFERENCES	112
APPENDIX.....	137

List of Tables

Table 2-1 An example of the decision table	31
Table 2-2 The 18 original driving factors	32
Table 2-3 An example of the discretized decision table	36
Table 2-4 An example of a decision table for extracting decision rules from RST.....	38
Table 3-1 Land-use classes identified in the generated land-use maps	50
Table 3-2 Selected factors using Rough Set Theory.....	55
Table 4-1 Parameters that need to be specified for defining a scenario	72
Table 5-1 Selected factors using Rough Set Theory feature selection	79
Table 5-2 $K_{Simulation}$ statistics for each land-use class for the simulation results obtained with the original 18 factors and the RST selected factors respectively	81
Table 5-3 $K_{Simulation}$ indices obtained for the patch-based and the cell-based CA models.....	88
Table 5-4 Landscape metrics (i.e., number of patches (NOP), mean patch size (MPS), and patch size coefficient of variation (PSCOV))	90
Table 5-5 Land composition for the projected land-use maps under the three scenarios in 2041.....	102
Table 5-6 Land consumption for each scenario over the simulation period.....	104

List of Figures

Figure 1-1 Illustration of Conway’s Game of Life simulation	8
Figure 2-1 Location of the study area (within the portion of the Elbow River watershed shown on the right of the dashed line)	25
Figure 2-2 Land-use map of the year 2006.....	26
Figure 2-3 Major methodological steps for factor selection using RST.....	30
Figure 2-4 Neighborhood configurations in the land-use CA model	34
Figure 2-5 Influence of the number of vegetation cells in NH2 on each type of land-use change	35
Figure 3-1 Study area (shown in the black box).....	46
Figure 3-2 Historical land-use map of the year 2011	51
Figure 3-3 Flowchart of the patch-based CA model.....	53
Figure 3-4 Probability map for the transition from Agriculture to Under Development (brighter tone showing higher probability values).....	60
Figure 4-1 Growth corridor for the eastern Elbow River watershed (IDP, 2011)	67
Figure 4-2 Land capacity for agriculture (Canada Land Inventory 2008).....	70
Figure 4-3 Flood hazard map of the Elbow River (Government of Alberta 2012)	71
Figure 5-1 Comparison between historical land-use map (a) and the results generated with the patch-based (b) and the cell-based CA models (c) for the year 2011	85
Figure 5-2 Comparison of the land-use class Under Development as observed in the historical land-use map and in the maps generated with the patch-based and the cell-based CA models for the year 2011	86
Figure 5-3 Simulated land-use map for the year 2041.....	92
Figure 5-4 Projected area for country residential over the simulation period	95
Figure 5-5 Projected area for urban residential over the simulation period	96
Figure 5-6 Projected area for agriculture over the simulation period.....	97
Figure 5-7 Projected area for forest over the simulation period	98
Figure 5-8 Projected land-use maps for the year 2041 under the three scenarios	101

List of Abbreviations

Symbol	Definition
CA	Cellular Automata
GIS	Geographic Information System
RS	Remote Sensing
RST	Rough Set Theory
NH0	Neighborhood #0
NH1	Neighborhood #1
NH2	Neighborhood #2
WOE	Weight of Evidence
U	Under Development
UR	Urban Residential
CR	Country Residential
A	Agriculture
F	Forest

Chapter One: Introduction

Land-use change has emerged as a key driver of environmental change with its potential effect of causing many unintended environmental and socioeconomic problems, such as the local impact on hydrology (Harbor 1994), the displacement of agriculture and forest (Kueppers et al. 2004), the decline in wetlands and wildlife habitats (Serneels and Lambin 2002), the degradation of ecosystem compositions, and the global impact of changes in atmospheric compositions (Foley et al. 2005). Understanding land-use changes can facilitate environmental sustainability by improving land management, enhancing the capability of assessing and predicting future land-use change trends, and advancing our knowledge of key land-use processes (Veldkamp and Lambin 2001). Recognizing the importance of land-use change, in 1994, Land Use and Cover Change (LUCC) was launched as a core project of the International Geosphere–Biosphere Program (IGBP) and the International Human Dimensions Program (IHDP) (Verburg 2006, Veldkamp 2009). The aim of the LUCC project was: 1) to develop fundamental understanding of the human and biophysical dynamic of land-use changes, 2) to develop robust models to predict and project land-use/cover change, and 3) to develop an understanding of land-use/cover dynamics through systematic and integrated case studies. With the increasing awareness of the importance of land-use/cover change, a novel discipline of land use science has emerged (Gutman et al. 2004).

1.1 Models for studying land-use change

Land-use change models are essential tools for understanding land-use dynamics, examining land-use causes, analyzing alternative land-use change consequences, and supporting land-use

planning and decision making (Lambin et al. 2000 Verburg et al. 2004, Li 2011, Silva and Wu 2012). Over the last century, a vast variety of land-use models have been developed.

Before the 1950s, most land-use models were based on spatial economic theory. One of the oldest contributions was Von Thünen's land rent theory of concentric rings dating back to 1826. According to this theory, land close to the city centre is used most intensively and the value of the land decreases outwards (Perraton and Baxter 1974). Other approaches that were proposed during this period include Weber's classical triangle of industrial location (1909) and Lösch's theory of economic regions (1940).

Following the advances in computational facilities, in the late 1950s, computer-based urban models arose. Land-use models of this period attempted to integrate the explicit representation of the time dimension (Crecine 1964, Paelinck 1970, Wegener et al. 1986). Microeconomic theory was the dominant foundation in land-use models. The microeconomic-theory-based land-use models focus on individual landowners who make land-use decisions with the objective to maximise expected returns gained from the land. The Lowry model is one of the best-known models of this type, which predicts the location of urban residential and basic industry (Lowry 1964). However, most land-use models developed from the 1950s to mid-1980s have the major limitation that no spatial dimension was taken into consideration.

From the end of 1980s, the spatial dimension was introduced into land-use models, which have resulted in spatially explicit land-use models becoming the dominant modelling framework. Land-use systems were then considered as complex systems composed of interacting individual

components (Batty et al. 1997). What emerges from these interactions cannot usually be predicted simply by analysing its components (Ashby 1956, Wolfram 1984). Complex systems are associated with many unique characteristics including feedback, non-linearity, adaptation, self-organization, path dependency, and emergence (Bennett and McGinnis 2008). Feedback is produced by the interaction between components, where the current state of the system will directly or indirectly influence its future state. It leads to non-linearity of the system where the relationship between the rate of change and the variables are not constant through time. The system adapts itself according to the interaction among its components and feedback processes. The system is path dependent, i.e., its current state depends on what has occurred in the past. Self-organized structures or patterns are formed caused by the interaction, feedback, and adaptation between the components within the complex adaptive system.

Along with the recognition of complex systems, land-use change models also evolved from a deterministic way of representing reality to complex adaptive systems. Scientists developed spatially explicit models to enable the simulation of land-use systems, among which four main categories can be identified that reflect the perspective of different disciplines: spatially-explicit econometric modelling, spatial allocation modelling, agent-based modelling, and Cellular Automata (CA).

Spatially-explicit econometric models simulate land-use change based on information about individual landowners' economic decision where a land parcel is encouraged to change to the land use that will maximize net returns over time based on its attributes and location (Irwin 2010). These models have a strong basis in economic theory and are used to analyze behavioral

responses to land-use policy changes. Some examples include: Chomitz and Gray's (1996) model of deforestation where the landowners maximize expected returns by determining the land-use type based on the rent; Fragkias and Geoghegan's (2010) model for simulating commercial and industrial land-use changes; and Suarez-Rubio et al.'s (2012) spatially-explicit econometric model to study exurban development near Washington, DC. A major limitation of these models is that they often require detailed economic data that are not readily available, such as individual home sale prices, lot characteristics, etc. (Wainger et al. 2007). Another criticism is that they are unable to capture the spatial contagion and repelling properties among land parcels (Irwin and Bockstael 2002).

Spatial allocation models tend to utilize neighborhood conditions to correlate certain types of land-use changes. Conversion probabilities at each location are calculated using transition rules that are based on characteristics of the location and the neighbourhood. A representative model of this group is CLUE, which simulates the geographical pattern of land-use change based on the local and regional suitability of the locations (Veldkamp and Fresco 1996, Verburg et al. 1999, Verburg and Veldkamp 2001, Verburg and Overmars 2009, Verbug et al. 2012). The suitability for different land-use types is determined by quantified relations using multiple regression models between land use and a large number of explanatory factors including biophysical and socio-economic factors. Land-use conversion is typically only based on suitability. These models are easy to understand and relatively easy to use. However, since they do not consider the interactions among the land use in the neighborhood, their major limitation is that the simulated land-use change does not have the ability of influencing the land-use changes in following time steps.

Agent-based models include a large and diverse class of simulation models characterized by interacting autonomous agents that have the ability to make decisions based on changing conditions (Parker et al. 2003). The basic unit in the agent-based models is the agent who represents a discrete entity having its own goals and the ability of adapting and modifying its behaviours. Agent-based models are good at simulating individual decision-making entities and the interactions among the agents. A well known agent-based model is FEARLUS where agents work as managers deciding the land use type and the according quantity (Polhill et al. 2008). However, these agent-based models usually attempt to integrate agents' potentially irrational behaviour, subjective choice, and complex psychology. These characteristics are difficult to quantify, calibrate, and sometimes justify, which complicates the model development and implementation.

CA are dynamic models used to investigate fundamental principles of system evolution and self-organization. They are a powerful tool for studying complex systems (of which land-use systems are examples) due to their ability to simulate dynamic spatial processes from a bottom-up perspective (Batty 2007, Iltanen 2012). These models are similar to the spatial allocation models in terms of using transition rules to govern land-use changes. However, CA-based models include spatial interactions between land uses in the neighbourhood and have proven to be particularly suitable for modelling the spatial dynamics of land-use systems due to their natural affinity for representing complex spatial forms (White and Engelen 1997, White and Engelen 2000, Jenerette and Wu 2001, Benenson and Torrens 2004, Verburg and Overmars 2009, Nijs and Pebesma 2010, Santé et al. 2010, Stanilov and Batty 2011, Silva and Wu 2012, Garcia et al.

2012). Currently, CA-based models are among the most popular ways to simulate land-use changes (Clarke and Gaydos 1998, Caruso et al. 2007, Long et al. 2012).

A key advantage of CA is their simplicity; they can generate spatial-temporal patterns of great complexity through a limited set of relatively simple rules (Batty 2007, Wu and Silva 2010, White et al. 2012). The models' outcomes are typically a set of simulated land-use maps, which can easily be visualized, quantified, and analyzed by users (Jantz et al. 2004). Moreover, information from other models, such as population growth models, can easily be integrated into a the CA model to constrain land-use changes. This makes CA models suitable for testing various "what-if" scenarios and policies (White et al. 1997, Almeida et al. 2008, Hasbani et al. 2011). In addition, these models can be linked to Geographic Information Systems (GIS) to utilize their functionality and directly use the raster-based spatial data derived from remote sensing platforms. The main concepts of CA and its application for land-use change modelling will be discussed in details in the following section.

1.2 Cellular Automata for land-use change modelling

1.2.1 Basic concepts of CA

CA were originally invented by Von Neumann in the mid-1940s to provide a formal framework for investigating the self-reproducing features of biological systems (Von Neumann 1963). Traditional CA is typically defined having five components (Wolfram 1984, Batty et al. 1999; White and Engelen 2000):

- 1) *a space* represented as a collection of homogeneous cells,
- 2) a set of possible *cell states*, which characterize the cells in the space,

- 3) a definition of the neighborhood of the cell,
- 4) a set of deterministic or stochastic *transition rules*, and
- 5) a sequence of *time steps*.

In a basic CA model, the states of the cells are updated synchronously at each discrete time step according to a set of transition rules that determine their evolution based on the states of the cells within their neighborhood and some external constraints. During CA modeling, complex global structures emerge through local neighborhood interactions. This simple model is self-reproducing and computationally universal, and is able to reproduce any recursive function (Benenson 2004).

A simple and well-known CA model is the Conway's Game of Life, a computer game simulating how life evolves, devised by the British mathematician John Horton Conway in 1970 (Gardner 1970, 1971). Conway's initial motivation was to design a simple set of rules to study the microscopic spatial dynamics of population (Berlekamp et al. 1982). In this model (Figure 1-1), space is represented as a two dimensional grid; each cell in the grid has one of the two possible states: alive or dead. The central cell has a Moore neighbourhood (i.e., eight neighbours around the central cell). The cells evolve according to a set of predefined transition rules: (1) any alive cells with fewer than two alive neighbors dies, (2) any alive cells with two or three alive neighbors lives on to the next generation, (3) any alive cells with more than three live neighbors dies, (4) any dead cell with exactly three live neighbors becomes a live cell. The rules are applied to all the cells in the grid simultaneously.

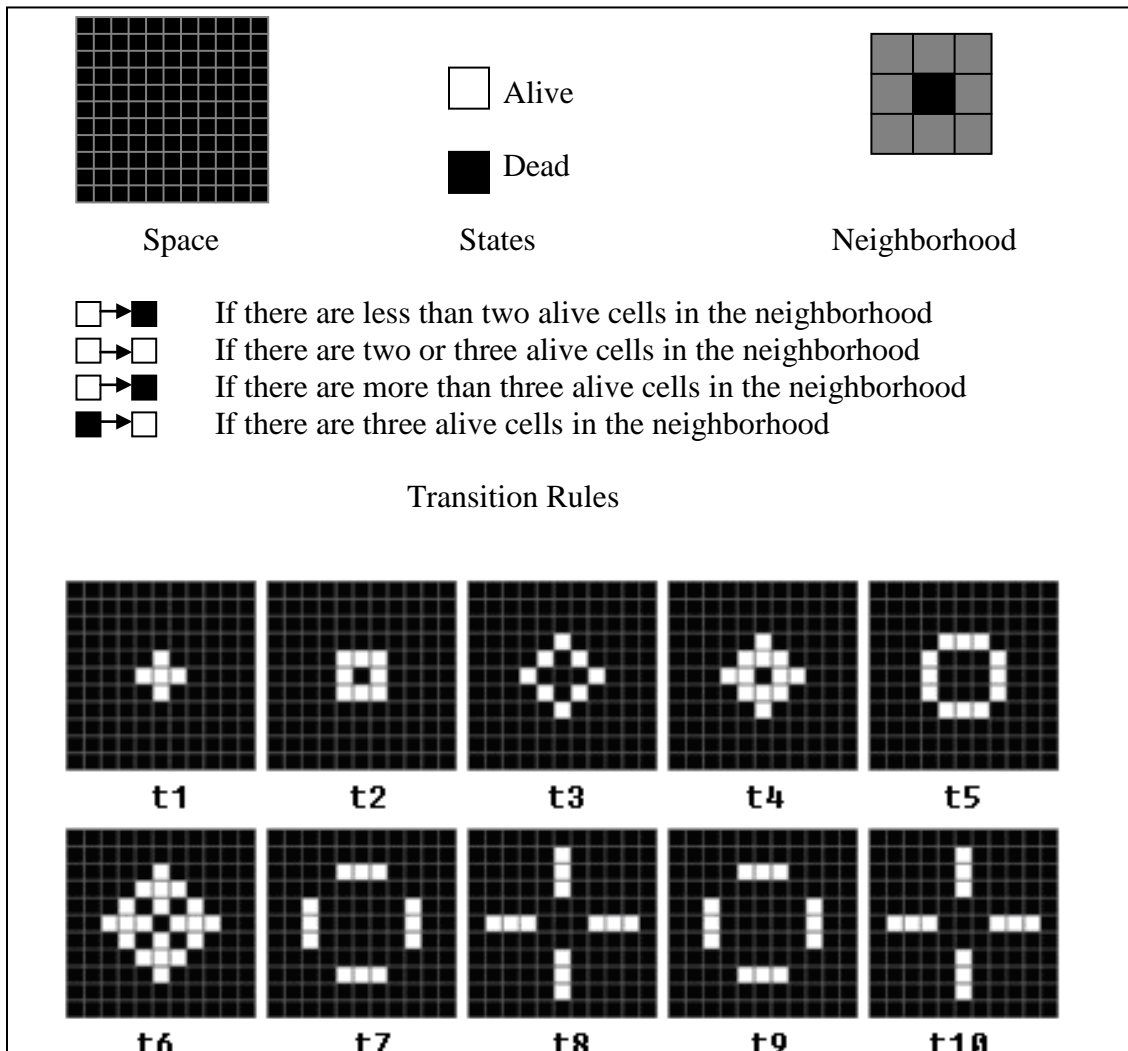


Figure 1-1 Illustration of Conway's Game of Life simulation

The Conway's Game of Life could generate a remarkable variety of patterns by following very simple rules, which has attracted much interest. This simple model introduced CA as an interdisciplinary tool for simulating complex systems and exploring their dynamics, which opened up a whole new world for the study of emergence and self-reproduction.

1.2.2 CA-based land-use models

CA have been widely used in simulating land-use systems ever since it was introduced by Tobler (1979) in geographic modeling. Tobler's ideas were adopted by Phipps (1989, 1992), who explored theoretical problems of CA in cluster formation. Couclelis (1985, 1989, 1997) investigated the complexity and structure formation in the context of spatial dynamics using CA, but did not apply them to specific cities. Later, various theoretical CA models were established. For example, Takeyama and Couclelis (1997) proposed a generalized modelling language, Geo-Algebra, to integrate CA modelling into a GIS framework. Batty and Xie (1994, 1999) and Xie (1996) developed a CA model for urban modelling which produced both urban spatial patterns and associated transportation networks. However, their models were not developed to model actual cities. Portugali and Benenson (1995) and Portugali et al. (1994) investigated general principles of urban self-organization using CA models.

In addition, the simple structure of the traditional CA has been extended to adequately represent real world land-use systems. For example, regular space in CA model has been modified using Voronoi polygons (Shi and Pang 2000, Hu and Li 2004), hexagonal grids (Gerling 1990, Eloranta 1997), and irregular vector objects representing real-world entities (Torrens and Benenson 2005, Stevens et al. 2007, Moreno et al. 2008, 2009, 2010). The neighborhood has been defined beyond the traditional von Neumann or Moore neighbourhood into extended neighbourhoods (White and Engelen 1997) and dynamic neighbourhoods (Moreno et al. 2009). Also, several techniques have been proposed to identify the transition rules from simple probabilistic and statistical methods (Wu 2002, Mitsova et al. 2011) to more sophisticated computational intelligence methods. For example, Li and Yeh (2001, 2002) used artificial neural

networks (ANN) for generating transition rules to simulate land-use changes in Dongguan, China. Liu and Phinn (2003) incorporated fuzzy-logic into their CA model to simulate land-use dynamics for the metropolitan region of Sydney, Australia. Feng and Liu (2012) employed a genetic algorithm to define transition rules in their CA model.

Benefiting from these theoretical explorations and structure improvements CA have become well-established tools for modeling land-use changes and have been applied to numerous real cities (Li and Yeh 2000, Santé et al. 2010, Li 2011). For example, White and Engelen (1997) and Engelen et al. (1997) developed an integrated CA land-use model that constrains land-use change with social-economic growth at a regional scale. Clarke (1997) developed a Cellular Automata Urban Growth Model (UGM), which was later renamed SLEUTH by incorporating Slope, Land use, Exclusion, Urban extent, Transportation, and Hillshade into the model. This model was then widely used for simulating land-use changes in various regions (Silva and Clarke 2002, Dietzel and Clarke 2008, Zhao et al. 2012, Kim and Batty 2012). Markov chains were also integrated with CA models (Myint and Wang 2006, Courage et al. 2009, Guan et al. 2011, Wang et al. 2012). In a Markov-CA model, the Markov chain process controls the temporal land-use dynamics while the CA model controls the spatial dynamics. DINAMICA (Soares-Filho et al. 2002) is another CA-based model that was originally designed for the simulation of deforestation and further applied to simulate urban land-use changes (Almeida et al. 2003, 2008, Pérez-Vega et al. 2012, Ferreira et al. 2012). These studies have demonstrated that CA are remarkably effective in generating realistic simulations of both land-use patterns and other spatial structures.

1.3 The rationale for studying land-use changes at fine spatial resolutions

Previous studies have adequately explored the usability of CA for land-use change simulation at spatial resolutions typically coarser than 20 m. Under those conditions, the models can contribute to evaluate land-use changes at a regional scale, which could facilitate the definition of appropriate environmental policies. However, land-use modeling at these resolutions are not capable of identifying subtle land-use changes (Houet et al. 2010).

CA models designed at fine spatial resolutions offer several unique features: 1) detailed land-use classes are easier to identify with clear characterization of their boundaries; 2) detailed land-use changes reflecting the socioeconomic attributes (e.g. country vs. urban residential) and ecological functions (i.e. green open space) can be detected at the land parcel scale (i.e. meters); 3) process-based land-use relationships (e.g. the influence of urban residential on agricultural land) are easier to discern at fine resolutions, which is essential to propose and implement local sustainable environmental policies and facilitate in situ decision making (GLP 2005, Pan et al. 2010, Zhao 2011, Stanilov and Batty 2011). As noted by Kok and Veldkamp (2001), these relationships are less apparent when the emergent land-use change patterns are observed at coarser resolutions where other processes, such as environmental or macro-economic factors, become dominant. Especially in a peri-urban area, which comprises an unbalanced mixture of urban and rural functions, the understanding of detailed land-use dynamics is of primary importance for land-use planning and decision making (Bryant et al. 1982, Antrop and Van Eetvelde 2000, Antrop 2004, Hoggart 2005, Meeus and Gulinck 2008). Planners and decision makers need to acquire precise information on land use and their changes over time in peri-urban

areas. To properly describe and simulate local land-use dynamics, fine-resolution datasets representing detailed land-use classes are required.

Some CA models have been applied at fine resolutions. For example, Tattoni et al. (2011) simulated forest dynamics using 5 m resolution data. However, land use was limited to only two categories: forest and non-forest. Also, Arai and Basuki (2010) explored the dynamics of mudflow disaster based on a CA at a resolution of 5 m. The study of Stanilov and Batty (2011) is one of the first to use high spatial resolution datasets (i.e., highly detailed historic Ordnance Survey maps at a resolution of 1:2,500 - 1.25 m resolution) to model the historic urban growth of London between 1875 and 2005. However, the fine-resolution datasets are only used for land-use classification to provide detailed land-use base maps. The simulation of their CA model was still based on the resampled dataset at a much coarser resolution of 25 m.

One reason for the lack of studies of land-use dynamics at fine resolutions is the challenge of consistent data availability (Bounfour and Lambin 1999, Xie et al. 2007). Recent advances in very high spatial resolution remote sensing imagery (spatial resolution less than 5 m) provide a great opportunity for acquiring fine-resolution land-use data at affordable prices. These images allow the identification of various categories of urban land uses such as high density and urban/country residential areas, commercial, industrial and institutional, as well as the capture of fine terrestrial features such as township roads, ponds, and small vegetation patches, which all provide great data sources for the calibration of land-use CA models at fine resolutions (Torrens 2006). However, even if fine-resolution datasets are now available, challenges still remain when

modeling land-use dynamics using CA with these datasets, which are discussed in the following section.

1.4 Challenges related to developing a land-use CA model at fine spatial resolution

There are several challenges associated with developing a fine-resolution CA model to understand land-use dynamics. In this thesis, two key challenges are addressed. The first one is associated with the appropriate method for dealing with the increasing number of factors and identifying the dominant ones that drive the landscape dynamics when calibrating the model. The second one involves defining a suitable space representation and simulation procedure in CA models for land-use modeling at fine spatial resolution.

1.4.1 Factor selection in CA modeling

With the increase of spatial resolutions, more detailed land-use classes are identified. These land-use classes and their interactions within a neighborhood are considered as internal factors in CA models. Moreover, several external factors also influence land-use changes including the physical characteristics of the area, the socio-economic conditions, infrastructure supply, demography, planning constraints, and environmental regulations (Liu and Phinn 2003). Land-use CA models often combine the influence of internal and external factors to define transition rules, which is typically done by calculating the transition probability of each cell at each time step of the simulation. However, some of these factors might be redundant while others might contain no information relevant to the land-use transitions (Fang et al. 2005). Incorporating all the internal and external factors in CA models, especially at fine resolutions, will undoubtedly result in complicated transition rules and sometimes make the model lose its generalization

potential. Another disadvantage of including a large number of factors is that the calibration and simulation time will also increase. Therefore, a critical challenge in defining a fine-resolution CA is to integrate the detailed land-use classes while keeping the transition rules as simple as possible. To do so, an important component consists of adequately selecting the key driving factors for CA model calibration.

However, there is still no general agreement on how to proceed for an optimal selection of the driving factors for the calibration of CA models. Most factor selection is usually based on modelers' knowledge regarding the land-use change processes (Li and Yeh 2004). Several methods for reducing factors exist which can be divided into two main categories: feature extraction and feature selection (Swiniarski and Skowron 2003). Feature extraction is the process of creating new features that contain most useful information by irreversibly transforming the original features (Liu and Motoda 1998). In a broad sense, feature extraction can be defined as a low-dimensional data representation. The advantage of feature extraction is that data become more computationally manageable. Presently, feature extraction is used in several studies to reduce the dimension for the calibration of CA models. Sui and Zeng (2001) acquired empirical evidence of the importance of different factors from historical land-use maps and use multiple regression to derive the weights of these factors used in the transition rules. Li and Yeh (2002) discussed the issue of correlation among spatial factors in urban simulation and applied principal component analysis (PCA) to extract a small number of new factors retaining most of the information contained in the original ones. PCA reduces the dimension of the data by analyzing the covariance between factors and identifying the principal components that contain the maximum variation in the data. However, the interpretation of the resulting components with

respect to spatial processes is not always straightforward. Lau and Kam (2005) performed a multivariate analysis of variance, combined with a multiple discriminant analysis to select the statistically meaningful factors in their urban land-use model. The standardized factors were linearly combined into discriminant functions. However, the number of factors in their model remained the same.

In contrast, feature selection is the process of removing features that are unnecessary or unimportant and ascertaining an optimal subset of features from an original dataset according to some criterion. The stepwise factor selection method used by Wu et al. (2009) in their CA model for simulating the dynamics of Chinese tamarisk is an example of feature selection. This method allows the identification of statistically significant factors that predict cell state transitions in CA model. However, the evaluation of each group of stepwise factors requires a pre-defined model, that is, the significance of a factor is assessed by evaluating whether there are significant changes in the model results when a factor is added or removed. Empirical methods are also applied to derive the factors. For example, Hagoort et al. (2008) combined literature search and expert interviews when selecting and adjusting the factors used for the calibration of their CA model. A total of 30 land-use experts were interviewed to identify the important driving factors and the according transition rules.

Feature selection can also be achieved by data mining and knowledge discovery methods, among which Rough Set Theory (RST) is an appealing method. RST has been proposed by Pawlak (1982, 1991, 1998) to unravel an optimal set of features and rules from an information system. Some researchers have demonstrated its effectiveness for feature selection in various fields.

Swiniarski and Skowron (2003) showed that RST can significantly reduce the features required for face recognition. Questier et al. (2002) successfully used RST to select features for an unsupervised clustering of different kinds of bacteria. Ahmad et al. (2007) applied RST feature selection to identify important factors for measuring companies' performance from a financial dataset. However, very few studies have been conducted that assess the potential of RST in the context of CA modeling. Hassen and Tazaki (2003) used RST to generate transition rules of a CA model to simulate traffic systems. Yang and Li (2006) integrated RST, a CA model and a geographic information system (GIS) to simulate the complex land-use patterns in Shenzhen city, in China. These studies revealed that RST can provide adequate transition rules for the calibration of CA models.

1.4.2 Space representation in CA models

The second challenge of designing a fine-resolution CA model lies in finding a suitable space representation for the land-use CA models. Space representation refers to how geographic space is represented in CA models. Using the context of space representation, CA models can be divided into two categories: 1) traditional cell-based CA where space is represented as a collection of regular cells (White and Engelen 2000); 2) entity (object)-based CA model, where space is defined as a collection of objects with irregular shape and size corresponding to the real-world entities (Moreno et al. 2009).

The traditional cell-based land-use CA models have a simple format and are capable of recreating the dynamics of complex land-use systems with high computational efficiency. These models are also compatible with the land-use maps generated from remote sensing images and

allows an easy integration with a GIS environment (Jenerette and Wu 2001, Dietzel and Clarke 2007, Wu and Silva 2010, Li 2011, Guan et al. 2011). In the traditional cell-based CA, one important assumption is that the cells are considered as interdependent entities (Deal and Schunk 2004). This implies that each cell represents a homogenous land-use entity that evolves individually as influenced by its neighbourhood; this means that the cell size in the traditional CA model should be close to that of the land-use features that evolve in the real world (Kok et al. 2001, Samat 2006). For example, at the spatial resolution of 60 m a cell may represent a residential area whose state can change over time. However, at fine resolutions (5 m), this assumption is not reasonable. On the fine-resolution land-use maps, the size of the cells is typically smaller than the common land-use entities being simulated (a cell might represent only a component of the residential area such as a street or a front yard). In that situation, the evolution of the land use occurs in an aggregated way that should be represented as a cluster of adjacent cells changing simultaneously during the simulation. For example, several cells that correspond to an area of high density residential must be updated as a group to represent a single spatial entity rather than individual cells having different attributes and being sparsely distributed over that area. Therefore, at fine spatial resolutions, the traditional cell-based CA is challenging and has difficulty in generating reliable land-use patterns by introducing *salt-and-pepper* noise (Wu 2002).

Many researchers argue that the CA models should rely on real lots and parcels - the land-use features that planners and developers consider when making decisions (Benenson and Torrens 2004, Blaschke 2006, Samat 2006, Schmit et al. 2006). Therefore, an improvement to the traditional cell-based CA is an entity-related (or object-based) CA, where real-world entities with

explicit geometry are directly represented in the model. Landis (1994) developed an urban growth model called California Urban Futures (CUF) based on the Developable Land Units (DLU). DLUs are represented as polygons that are constructed through the union and/or intersection of different environmental, market, and policy map layers. Hu and Li (2004) developed an object-based CA in which geographical entities were represented as points, lines, or polygons according to their real shapes and size. In favor of planning and decision making, Stevens et al. (2007) and Stevens and Dragicevic (2007) developed iCity in which an urban area is partitioned into discrete land-use units based on cadastral information and represented as a collection of polygons. Similarly, Pinto and Antunes (2010) developed an irregular CA based on census blocks to determine the land-use demand by considering the evolution of population and employment densities over time. This model enabled an easy connection between land use and demographic and socio-economic information. Also, Onsted and Clarke (2011) incorporated zones derived from policies such as the Farmland Security Zone (under California's Williamson Act). Zones were represented using polygons managed by the SLEUTH model for simulating urban growth and land-use changes in Tulare County, California. However, while considerably improving the representation of real-world entities within CA, these models still lack the ability to simulate changes of entities' geometry.

Addressing this challenge, Cao (2011) developed a vector-based CA model which allows the change of the geographic objects' geometry. However, the neighborhood is defined topologically, as a buffer around each object. Further improvement was presented by Moreno et al. (2008, 2009, 2010) who developed an object-based CA model that allows the change of geometry of each object and in which a dynamic neighborhood defined semantically is

implemented; this neighborhood includes both the adjacent objects, and the objects separated by other objects whose states are favorable to the transitions to the state of the central object. These models generate a more realistic representation of land-use change when compared to a traditional cell-based CA model.

A comprehensive framework to better capture the complexity of urban system dynamics referred to as Geographic Automata Systems (GAS) was proposed by Benenson et al. (2002) and Torrens and Benenson (2005). In GAS, a city is described as having two interacting layers. The first layer consists of immobile urban components represented as fixed objects using polygons or polylines (buildings, land parcels, roads, and parks) simulated by a CA model. The second layer contains the mobile urban components represented as non-fixed objects (urban decision makers, pedestrians, vehicles, and householders) described in an agent-based model. This original framework was further improved to allow the change of geometry of the objects (Hammam et al. 2007, Moore 2011).

In the entity-based CA models, each polygon is assumed homogeneous and represents one land-use or land-cover feature. However, this homogeneity assumption may be broken when the spatial resolution becomes finer and additional details of the geographical entities appear. For example, an agriculture area cannot be seen as a homogeneous object anymore when large variations exist inside the area, such as the slope, accessibility, and suitability. At fine resolutions, it may not be realistic to apply a single transition probability value to the entire land-use entity, as changes may happen only in a portion (not necessarily located at the boundary) of the entity. This requires that the homogeneous object space of the CA is replaced by

geographical space where the internal heterogeneity of the entities is explicitly considered. Therefore, at fine resolution, a more suitable space representation should be used in the CA model, where realistic land evolution is simulated. This will not only avoid the discrete conversion of individual cells in the traditional CA models, but also take into consideration the heterogeneity inside each object in the entity-based CA models (Wu 2002).

1.5 Objective of the thesis

The primary goal of this research is to develop an innovative CA model to simulate land-use changes at fine spatial resolutions, for the purpose of advancing our understanding of detailed land-use changes and their driving factors. To achieve this goal, two critical challenges previously discussed in Section 1.4 are addressed, with the following three key components to be completed:

- 1) To propose and implement an effective way to identify dominant driving factors required for the calibration of land-use CA models. This thesis proposes a novel method using RST (Rough Set Theory) for selecting key land-use driving factors.
- 2) To develop an appropriate representation of space for the fine resolution CA model. An innovative approach is proposed to simulate land-use changes using a CA model at fine resolution (5 m) in which the real-world entities are represented as patches. A patch refers to a collection of adjacent cells that, when combined together, represent an entity differing from its surroundings in nature or appearance (Weins 1976, Fu and Chen 2000).

- 3) To simulate land-use changes in the future (from 2011 to 2041) under three different planning scenarios using the developed patch-based CA model, namely, *business-as-usual scenario* that evaluates the spatial consequences under the same condition obtained from the historical land-use maps, *protective growth scenario* that incorporates the watershed and agriculture protection, and *smart growth scenario* that encourages the growth of high-density land use and at the same time takes into consideration the projected population growth in the study area.

1.6 Assumptions and limitations

This research has been done based on the following assumptions:

- 1) The selection of the factors used for the CA models defined in this thesis is based on both data availability and literature review. We assume that these factors are sufficient in representing the land-use dynamics in the study area.
- 2) The road and water network impose important influence to the CA models developed in this study. However, since this research represents the first step of exploring land-use CA modelling at fine resolution, to simplify the modelling process, the road and water network were assumed unchanged during the modelling period.
- 3) The patch-based CA model utilizes mean patch size and minimum patch size information extracted from historical data when simulating land-use changes. We assumed that this information remains unchanged during the simulation period.

1.7 Organization of the thesis

The remaining part of the thesis is organized as follows:

Chapter 2 describes the methodological framework based on a data mining technique, RST, to guide factor selection for the calibration of CA models. Since the number of driving factors in fine-resolution CA models is usually large, to make the development and evaluation of the RST factor selection framework easier, RST will be first tested in an area using a dataset with a reduced number of land-use classes (5 classes) and original driving factors (18 factors) at coarse resolution (30 m).

In Chapter 3, a patch-based CA model is introduced to simulate land-use changes at 5 m resolution. The methodological procedures required for developing the patch-based CA are described. The RST framework developed in Chapter 2 was applied to a fine resolution (5 m) dataset with a larger number of land-use classes (14) and driving factors (47 factors).

In Chapter 4, three scenarios designed according to the information extracted from the local municipal development plans are described: the *business-as-usual scenario*, the *protective growth scenario*, and the *smart growth scenario*. These scenarios were simulated using the patch-based CA model for a period of 30 years from 2011-2041 with a time step of three years.

In Chapter 5, the results for RST factor selection described in Chapter 2 are presented and simulation results generated using the RST factors are compared with the ones generated using all the factors. The performance of the patch-based CA model described in Chapter 3 is evaluated by comparing the results generated by a traditional cell-based CA. Finally, the impact

of the three scenarios on non-developed land consumption are evaluated and compared to each other.

In Chapter 6, the general conclusions from this research and recommendations for future work are presented.

Chapter Two: Identifying Dominant Driving Factors using Rough Set Theory

This chapter describes the methodological framework of using Rough Set Theory (RST) to guide the factor selection for the calibration of a CA model. This data mining approach was tested to simulate land-use changes at a resolution of 30 m in a portion of the Elbow River watershed adjacent to the City of Calgary, in southern Alberta, Canada. This technique is specifically useful in the context where a large number of factors are considered. The application of the RST for factor selection in a fine-resolution CA will be described in Chapter three.

2.1 Study area and datasets

The study area is the eastern portion of the Elbow River watershed, located in southern Alberta, Canada, that covers an area of about 600 km² (Figure 2-1). The area is under considerable pressure for urban development due to Alberta's strong economy. About 5% of the watershed is within the City of Calgary, a fast growing city of one million inhabitants; 10% is located within the Tsuu T'ina nation, 20% within the municipal district of Rocky View No. 44, and the remaining 65% within the Kananaskis Improvement District. About 10% of the study area is covered by about built-up areas located mainly in the north-east part, while forest and other vegetation (agriculture, grassland and rangeland) represent approximately 30% and 35% respectively. Other areas are characterized by water and Tsuu T'ina land. The Tsuu T'ina lands are First Nations reserve lands which have undergone minimal changes over the last eight years will not be simulated here.

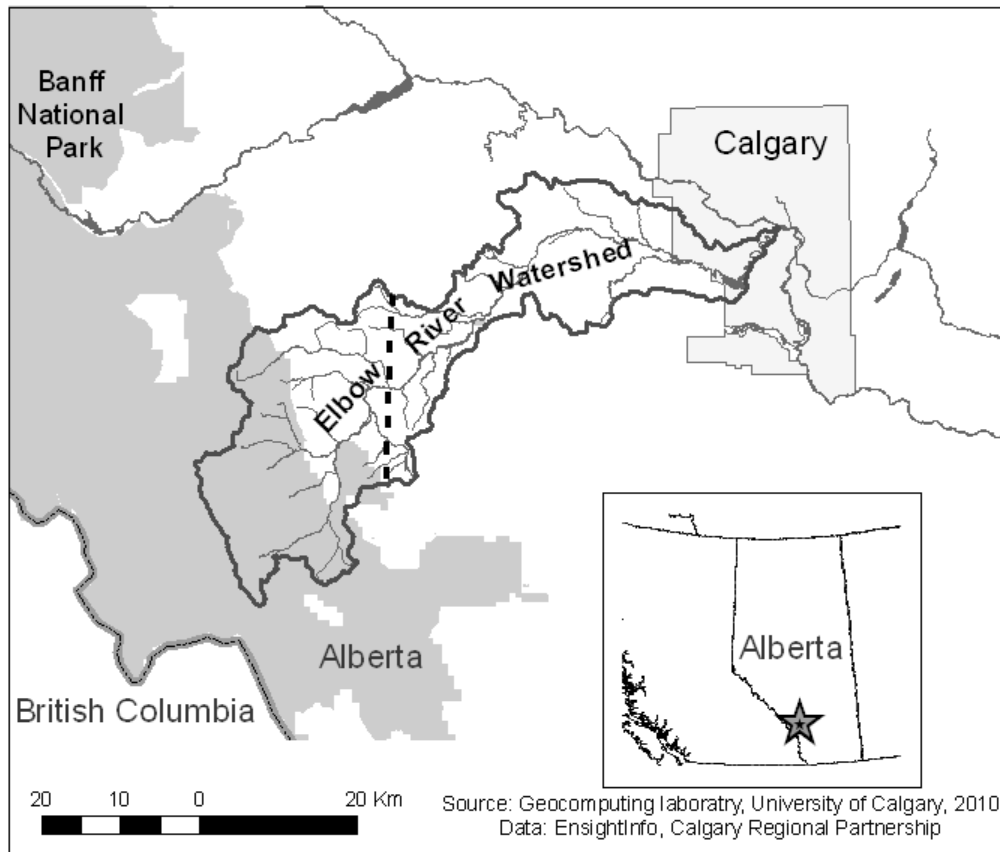


Figure 2-1 Location of the study area (within the portion of the Elbow River watershed shown on the right of the dashed line)

The historical land-use maps used in this study are the same set employed by Hasbani et al. (2011). They were generated from the classification of Landsat Thematic Mapper scenes, acquired during the summers of 1985, 1992, 1996, 2001 and 2006 at the spatial resolution of 30 m. Five dominant land-use/land-cover classes of the study area were identified: forest, built-up, water, Tsuu T'ina land, and vegetation which includes agriculture, grassland, and rangeland (Figure 2-2). The classified land-use maps were assessed through field verification in the summer of 2006. The overall accuracy is 0.79 for the year 1985, 0.77 for 1992, 0.80 for 1996, 0.82 for 2001 and 0.82 for 2006. Based on a sensitivity analysis conducted by Hasbani et al.

(2011), these land-use maps were re-sampled at the resolution of 60 m to reduce the computation time required for the calibration and simulation while maintaining the desired level of spatial details for the study.

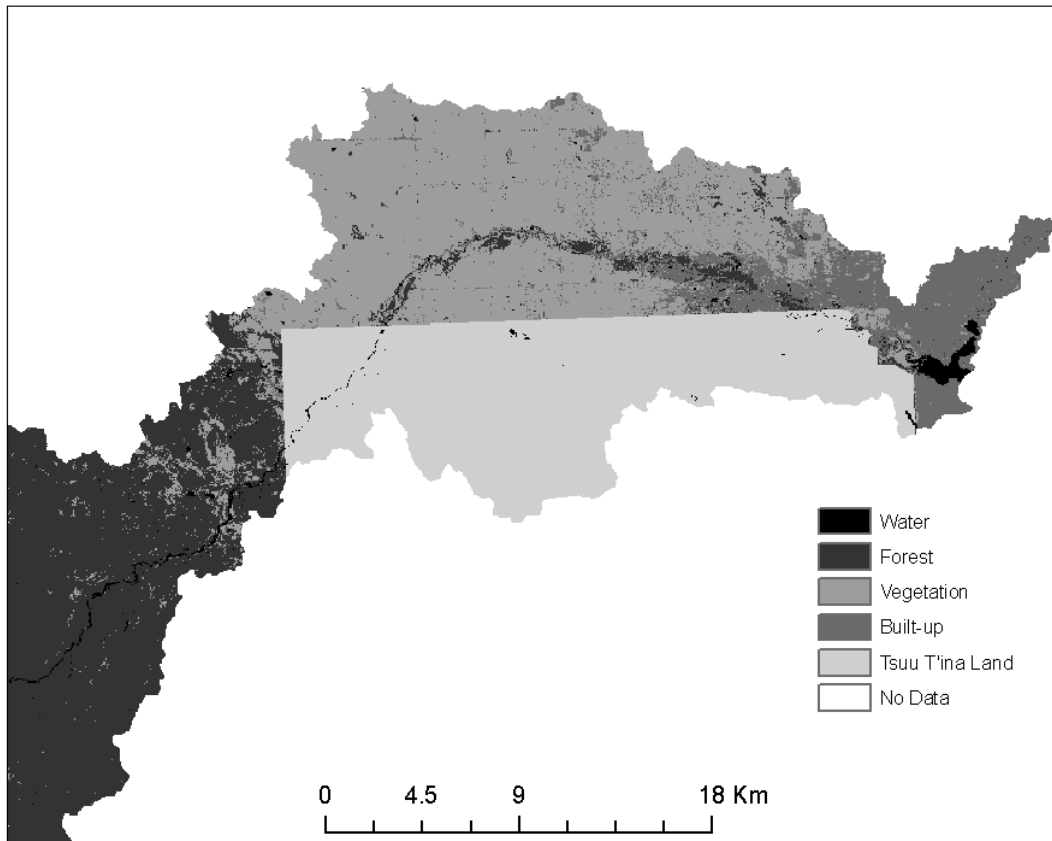


Figure 2-2 Land-use map of the year 2006

2.2 Fundamentals of Rough Set Theory for feature selection

In RST, information is defined as an approximation space, which can be represented as $T = (U, A)$, where U and A are two finite, non-empty sets. U is a certain set of objects called the universe, and A is a set of attributes or features associated with the objects (Pawlak 1991). Each attribute or feature $a \in A$ is associated with a set of values (V_a), called the domain of a . The attributes set,

A, can be partitioned into two subsets C and D, which represent conditional attributes and decisional attributes respectively.

If $P \subseteq A$, the indiscernibility relation $IND(P)$ can be defined as:

$$IND(P) = \{(x, y) \in U \times U : \forall a \in P, a(x) = a(y)\} \quad (2-1)$$

where x and y are two objects in the universe (i.e., $x, y \in U$), and $a(x)$ and $a(y)$ are the values of x and y of attribute a . If $(x, y) \in IND(P)$, x and y are considered to be indiscernible using the attribute set P (i.e. x and y cannot be identified or separated clearly).

All equivalence classes of $IND(P)$ in the universe can be defined as $U / IND(P)$. Each element in $U / IND(P)$ is a set of indiscernible objects of P . $U / IND(C)$ and $U / IND(D)$ correspond to the conditional classes and decisional classes.

For each universe subset, $X \subseteq U$, and attribute subset, $R \subseteq A$, X can be described using two sets: the lower approximation R_* , and upper approximation R^* , using the knowledge for R . The lower approximation of X is defined as:

$$R_*(X) = \bigcup \{E \in U \mid IND(R) : E \subseteq X\} \quad (2-2)$$

The upper approximation of X is defined as:

$$R^*(X) = \bigcup \{E \in U \mid IND(R) : E \cap X \neq \emptyset\} \quad (2-3)$$

The lower approximation contains all the elements that certainly belong to X , while the upper approximation contains the elements that possibly belong to X . If $R_*(X) = R^*(X)$, X is R -definable. Otherwise, X is a rough set with respect to R .

Different set of features determine different discernibility. There may be some feature in the information system which are redundant or irrelevant. According to Jensen and Shen (2007), a feature is considered irrelevant if it has no contribution to the decision feature. A feature is said to be redundant if it has high correlation with other features. RST is a method that can eliminate the redundant and irrelevant features and to extract features that are important to the information system. In RST, there are two important concepts that relate to this point. One is called *reduct*, which is a subset of features that can, by itself, fully characterize the knowledge in the decision table as the original set of features does. The concept of *reduct* is formally defined as: given an information system, a *reduct* is a minimal set of features, $B \subseteq A$, that gives the same indiscernibility relation as the original set of features: $IND(A) = IND(B)$. Feature selection using RST generates the minimal set of *reducts* that represent the same approximate space as the one that includes the whole set of features. Since there may exist several *reducts* for one single decision table, another concept in RST is *core*, which is the intersection of all *reducts*, which is defined as:

$$CORE_D(C) = \bigcap RED_D(C) \quad (2-4)$$

where, $RED_D(C)$ is the D -*reduct* of C (i.e. *reduct* from C for the information system D).

Core consists of the most important features that cannot be removed from a decision table without causing the collapse of the system (Pawlak 1998). This means that all attributes in the *core* are indispensable. The theory of Rough Set provides a technique to analyze data dependencies and identify fundamental factors from data.

2.3 Procedure for identifying the dominant driving factors using RST

In this section, the procedures of identifying the most important factors of the decision table using RST are described. The selected factors are then translated into transition rules used for the calibration of the land-use CA model. The methodology includes four main steps: (i) decision table preparation, (ii) selection of the driving factors using RST, (iii) generation and implementation of the CA model transition rules, and (iv) model simulation and results comparison (Figure 2-3).

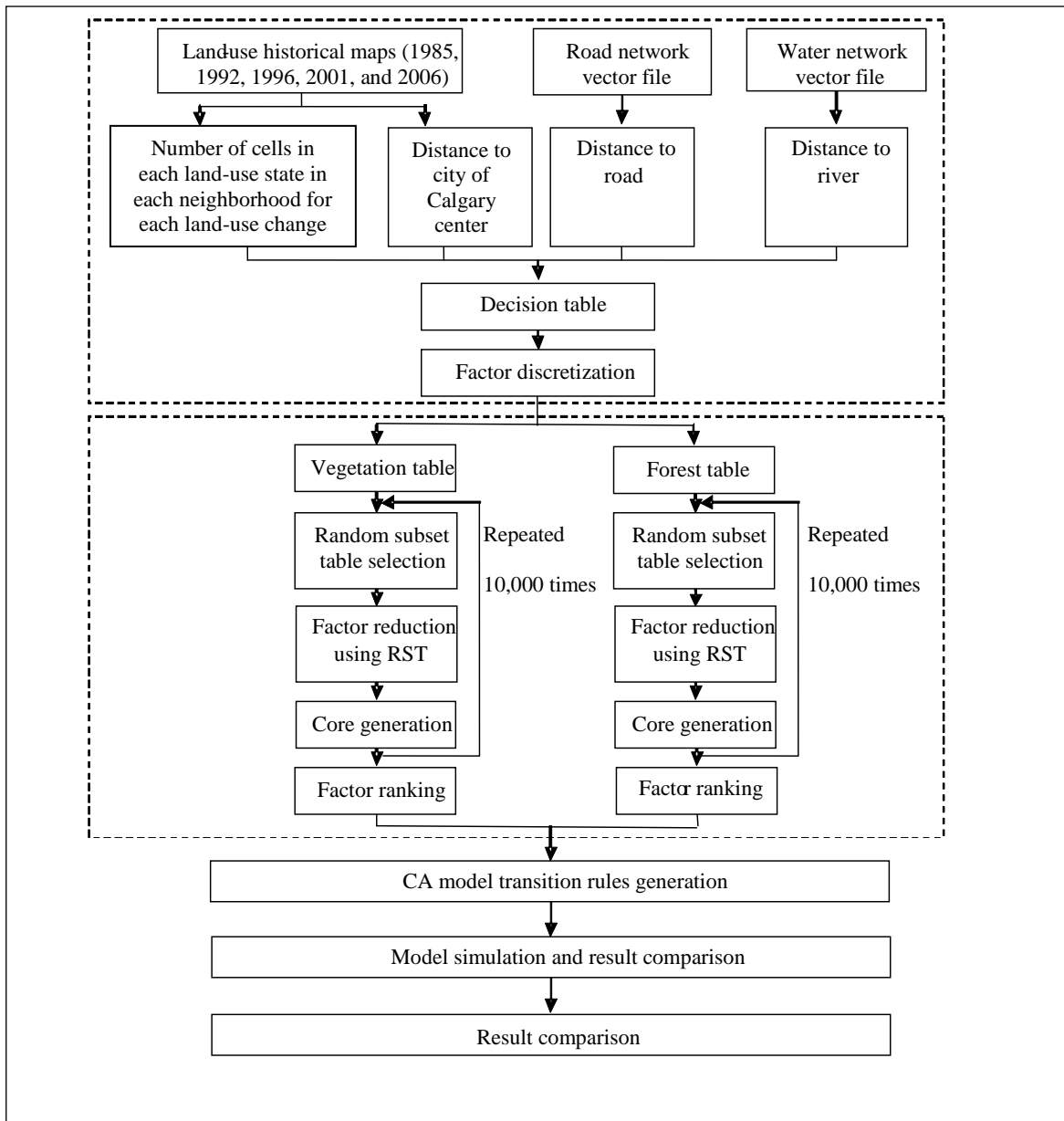


Figure 2-3 Major methodological steps for factor selection using RST

2.3.1 Decision table preparation

The decision table provides a standard data format for RST data processing (factor reduction). All types of original data (land-use maps) must be represented in a decision table. Each row in the table represents an individual record (the land-use change of one cell). Each column

represents some attributes of the records composed of conditional attributes (the initial land-use state, and the attributes corresponding to the factors associated to the land-use changes) and decision attributes (the final land-use state). An example of the format of the decision table is provided in Table 2-1.

Table 2-1 An example of the decision table

Record	Initial land-use state	Distance to a road (m)	Number of vegetation cells in neighborhood	... Final land-use state
1	Forest	1358	30	... Vegetation
2	Forest	2543	155	... Vegetation
3	Vegetation	549	220	... Built-up area

To generate the decision table, two steps were performed: (i) extracting the land-use changes (the records), and (ii) identifying the original land-use change driving factors and their corresponding values for each record.

Land-use changes were extracted from the comparison of the historical land-use maps for each time interval: 1985-1992, 1992-1996, 1996-2001, and 2001-2006. Four types of land-use changes were considered: 1) from Vegetation to Forest, 2) from Vegetation to Built-up, 3) from Forest to Vegetation, and 4) from Forest to Built-up. Changes from Built-up to other land-use classes were not simulated, since once a land parcel is developed, it usually remains in that state for a long period of time. Changes from and to water and changes within the Tsuu T'ina Land were minimal and not simulated in this study.

Table 2-2 The 18 original driving factors

Factors	
1	Distance to main roads
2	Distance to Calgary city center
3	Distance to a main river
4	Number of water cells in NH0
5	Number of forest cells in NH0
6	Number of vegetation cells in NH0
7	Number of built-up cells in NH0
8	Number of Tsuu T'ina Nation land cells in NH0
9	Number of water cells in NH1
10	Number of forest cells in NH1
11	Number of vegetation cells in NH1
12	Number of built-up cells in NH1
13	Number of Tsuu T'ina land cells in NH1
14	Number of water cells in NH2
15	Number of forest cells in NH2
16	Number of vegetation cells in NH2
17	Number of built-up cells in NH2
18	Number of Tsuu T'ina Nation land cells in NH2

Two types of driving factors were considered in this study based on the literature (White and Engelen 2000, Al-Ahmadi et al. 2009, Hasbani et al. 2011) and data availability. According to the characteristics of CA modeling, the interaction of a cell with the cells in its neighborhood is considered the dominant reason for land-use changes. Also, it is widely recognized that land use at one site is dependent on the land use at the neighboring locations (Tobler 1979); subsequently, neighborhood interactions are the key driving factors in land-use CA models (Verburg et al. 2004). Therefore, the first group of driving factors is the number of cells in each land-use state within the neighborhood of a central cell, which is referred as an internal driving factor.

In addition, according to many researches in urban theory, accessibility is considered fundamental to land-use changes (Stanilov 2003). To represent the influence of accessibility in this study, three external factors were utilized, namely: distance to main roads, distance to a main river, and distance to Calgary City center. The road and water network data were acquired to generate the first two factors. The factor distance to Calgary city center was measured using a digitized point of downtown Calgary. These distances were calculated for every cell at each historical year in the study area using the Euclidian distance function available in ArcGIS 9.2 (ESRI, 2007). The assumption here is that the road and water network remain unchanged during the simulation years. While additional factors might also be associated to land-use changes in the area, the selected ones are considered of significant relevance and in sufficient number for testing the applicability of RST for factor reduction.

To account for the influence from the cells of different distances, three neighborhoods were considered, called NH0, NH1 and NH2 respectively (Figure 2-4). Based on a sensitivity analysis previously conducted by Hasbani et al. (2011), neighborhoods corresponding to 3, 5 and 15 cells of radius from the center cell were selected. These neighborhood rings are all exclusive, that is, a cell can only be located within one single ring (defined by a radius), and there is no gap between two rings; this configuration generates a doughnut-shaped neighborhood. The influence of the neighbors to the central cells is considered constant within each neighborhood, while different between neighborhoods. This provides a total of 18 factors (Table 2-2). Each record of land-use changes from the historical data and the values of these 18 corresponding factors were combined to build the decision table.

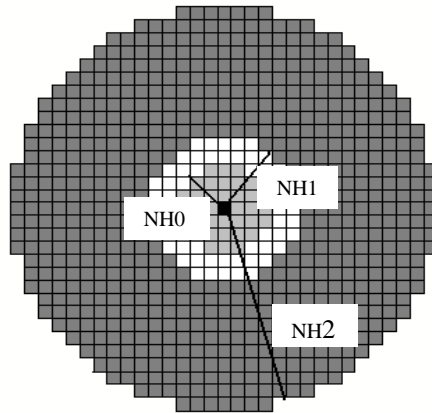


Figure 2-4 Neighborhood configurations in the land-use CA model

The values corresponding to the driving factors compiled in the decision table are continuous. However, to conduct the factor reduction with RST, these values must be discretized. Discretization is the process of partitioning attribute values into intervals and unifying the values over each interval (Nguyen 1998). For each factor, close values (e.g., the presence of 55 or 56 Forest cells in the neighborhood of a Vegetation cell) might have a similar influence on land-use change and could be grouped. To do so, each type of land-use change (e.g., from Forest to Built-up) and the corresponding values of the driving factors were combined in a table. Then the frequency of the change for each driving factor and each type of land-use change in the table was plotted as a histogram to depict the influence of each factor associated with each type of land-use change. If the frequency of change is similar for two values of a factor, then it means that the two values of this factor have a similar influence on the land-use change. Therefore these two values can be discretized within the same category.

Figure 2-5 represents the frequency of change histogram displaying how the factor ‘Number of Vegetation cells in NH2’ is associated with each type of land-use change. It reveals that some

numbers of Vegetation cells in NH2 have different influences on each type of land-use change, while some other numbers have a similar influence. For example, 450 and 470 Vegetation cells in NH2 have a similar influence on all the land-use changes. Since similar frequencies indicate similar influences of the factor on the land-use change, on these histograms, ranges of similar frequency of change for each factor were identified and grouped. This procedure is called factor discretization. A frequency of change histogram was generated for each driving factor. Then each driving factor was discretized into four different ranges of values based on the examination of the histograms of all the factors. At the end of this process, a discretized decision table was created including the initial land-use states, the final land-use states, and the discretized driving factors. An example of the discretized decision table is shown in Table 2-3.

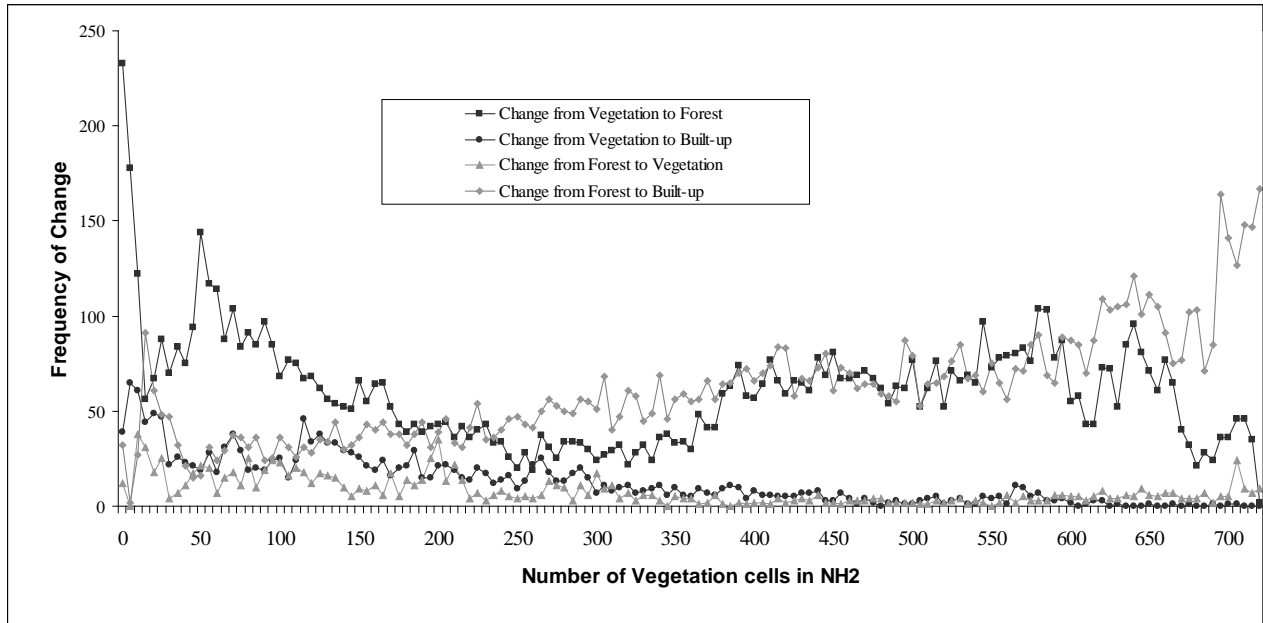


Figure 2-5 Influence of the number of vegetation cells in NH2 on each type of land-use change

Table 2-3 An example of the discretized decision table

Record	Initial land-use state	Distance to a road (m)	Number of vegetation cells in NH2	...	Final land-use state
1	Forest	(600, 1600]	(10, 130]	...	Vegetation
2	Forest	(1600, 4500]	(130, 260]	...	Vegetation
3	Vegetation	(0, 500]	(130, 260]	...	Built-up area

2.3.2 Selection of the driving factors using RST

To select the driving factors for the calibration of the land-use CA model, the discretized decision table was processed using Rosetta, a Rough Set toolkit dedicated to data analysis (Øhrn et al. 1998, Øhrn 1999, Rosetta development team 2009). This software contains all the commonly used Rough Set algorithms and provides an easy-to-use GUI front end. A command version of the software is also available for batch processing. Three steps were performed: (i) reduce the original set of factors, (ii) find core sets for each decision table, and (iii) rank the cores to find the most important factors for calibration of the land-use CA model.

To determine the driving factors of land-use changes that have occurred from different initial land-use states, the decision table was divided into two sub-tables, namely the Vegetation and the Forest tables. Only Vegetation and Forest were considered because the land-use changes in the study area only refer to these two initial states. There were 10,623 records of changes in the Forest table (corresponding to changes from Forest to Built-up and to Vegetation), and 10,121 records of changes in the Vegetation table (corresponding to changes from Vegetation to Built-up and to Forest). The Forest and Vegetation tables were imported into Rosetta to proceed with the factor reduction.

When dealing with real world problems, errors are inevitably created in spatial datasets during the data collection process. In our case, such errors may be caused by the presence of misclassified cells in the historical land-use maps, or the lack of precision in registration of the data to a particular coordinate system. RST, which is based on rigorous logic and mathematics, faces difficulties in factor reduction when such errors are introduced into the dataset (Han et al. 2001, Wang 2005), even if they are minor. Suppose an information system can be totally described by a minimal set of attributes. If some errors are introduced in the system attributes, it might be impossible to retrieve the minimal set of attributes. To address this problem, a random sampling approach was used. For both the Vegetation and Forest tables, a subset of records was randomly extracted and imported into Rosetta for the factor reduction using a genetic algorithm (Vinterbo and Øhrn 2000). This method iteratively evaluates a population of potential subsets of factors and determines the best subset to satisfy the fitness function. Several trials showed that a subset of 5% of the original dataset can effectively reduce the error in the original dataset affecting the factor reduction. Since there may be more than one reduct for one single decision table, core was generated as the intersection of all the reducts for that decision table. The factors constituting the core are the most important ones for the decision table. This step was repeated 10,000 times to ensure that all records were selected at least once. Therefore, 10,000 groups of cores were generated for the Vegetation and Forest tables respectively.

These factors were then ranked using a voting strategy. If a factor appeared once in the factor reduction, it was allocated one point. After 10,000 times, the factors having the highest points were considered to be the most important. A threshold of 5,000 points was empirically applied to select the number of factors considered as suitable for this study. The selection of this threshold

was made on the basis that if a factor appears more than 50% of the time in the reduction process then it is significant. This threshold also allowed an appropriate number of factors to be kept. Nine and eight factors were chosen for the Vegetation and the Forest land uses respectively.

2.3.3 Generating and implementing the CA model transition rules

For each type of land-use change, transition rules were extracted using Rosetta. The rule extraction procedure aims at identifying patterns in the decision tables and expressing them as decision rules as illustrated in Example 1.

Example 1:

Consider the dataset in Table 2-4 with four records, two factors (F1 and F2), and decision D. A rule extraction process transforms this dataset into a set of rules by binding the values responsible for the decision D:

Rule1: IF (F1=0) AND (F2=1) THEN (D=a).

Rule2: IF (F1=0) AND (F2=0) THEN (D=a XOR b). The probability of change to a or b is calculated from the historical data by counting the actual number of changes to both a and b.

Rule3: IF (F1=1) THEN (D=b).

Table 2-4 An example of a decision table for extracting decision rules from RST

Record	F1	F2	Decision
1	0	1	a
2	0	0	a XOR b
3	1	0	b
4	1	1	b

The transition rules for the land-use CA model were generated through binding values of the selected factors that are responsible for the land-use changes based on both the Vegetation and the Forest decision tables. These rules are represented as “IF factor values THEN land-use changes”. An example of a transition rule is displayed below:

IF “land-use initial state” = “Forest” AND “Number of Built-up cells in NH0” = (11, 20] AND “Number of Vegetation cells in NH1” = (0, 22] AND “Number of Built-up cells in NH1” = (21, 46] AND “Number of Forest cells in NH2” = (0,60] AND “Number of Vegetation cells in NH2” = [0, 50] AND “Number of Built-up cells in NH2” = (51, 250] “Distance to river” = (0, 600] AND “Distance to city center” = [0, 1500] THEN “land-use final state” = “Built-up”.

To complete the generation of the transition rules, the likelihood of application of each transition rule must be assessed. To do so, using the historical land-use maps, the number of cells updated by each transition rule is identified, counted and divided by the total number of cells associated with the corresponding land-use transition. An example of the rules generated using RST for the transition from Forest to Built-up is shown in Appendix 1. Given the large number of rules, only the rules with the highest frequencies of transition are presented.

2.3.4 Model simulation

Simulating future land-use changes involves the following steps:

- 1) loading the historical land-use maps and the transition rules generated from section 2.3.3;
- 2) extracting the number of cells associated with each type of land-use change and each transition rule;

- 3) for each transition rule, identifying the suitable cell values for each factor and changing the cell state according to the transition rule;
- 4) updating the transition rule statistics;
- 5) creating the resulting map;
- 6) updating the time step and repeating the process until the simulation is completed.

Once the historical land-use maps and the transition rules are loaded, the number of cells associated with each type of land-use change and each transition rule is extracted from the land-use maps. When the simulation starts, the total number of cells associated to each type of land-use change from the previous time step is acquired. Similarly, the likelihood of application of each transition rule is calculated by averaging the likelihood from the previous time steps as described in section 2.3.3. The likelihood represents the percentage of cells associated with the corresponding type of land-use change. To obtain the number of cells associated with each transition rule, the likelihood of the rule is multiplied by the average number of cells associated with the corresponding land-use change.

The next step is to identify the suitable cells of change for each transition rule and to update their land-use state. The key driving factors for each cell were acquired and classified using the data preparation for Rosetta. The suitable cells for each transition rule were identified using the values of the driving factors. Recursively, for each type of land-use change (e.g. Vegetation to Forest) and each transition rule, a random cell among the suitable ones is selected and its land-use state is updated until the proper number of cells has changed state or no more cells can change state. If there are not enough suitable cells for a transition rule, the remaining number is

dispatched among the other rules that are entitled to the same type of land-use change in order to have the correct number of cells associated with the type of change.

Then, the number of cells associated to each type of land-use change and the likelihood of each transition rule from the previous time step are updated, as the number of cells that change for each land-use transition and each rule for the following time step may be different. If the targeted number of cells, as extracted in step 2, is not exactly the same as required, the difference is recorded and subtracted or added to the number of cells at the next time step of the simulation. This feature breaks the linearity induced by the extraction method and adds flexibility to the simulation.

The resulting map at each time step of the simulation is created by merging the cells that updated their land-use states and the ones that did not. The output is a series of geo-referenced maps that can be integrated in GIS software for further analysis. Then, the time step is updated and if further simulation is required, the model repeats the process to the end of the simulation.

2.3.5 Results comparison

To evaluate the performance of RST for factor selection, two groups of transition rules generated using two categories of factors were applied to calibrate the CA model: i) the factors selected using RST and ii) the original 18 factors. Then two groups of simulation results were generated accordingly for the years 1992, 1996, 2001, and 2006. These simulation results were then compared to the historical land-use maps.

To evaluate the quality of the CA simulation results, two methods are typically used in the literature: aggregate similarity and spatial similarity (Brown et al. 2005). Aggregate similarity utilizes statistics to describe patterns in maps. For example, Clarke and Gaydos (1998) assessed their model performance with four statistical measures: a r -squared fit between the actual and predicted number of urban cells, a r -squared fit between the actual and predicted number of edges, a r -squared fit between the actual and predicted number of separate clusters in the urban distribution, and a modified Lee-Sallee shape index. Another solution was provided by Wu (1998) who applied the coefficients of the density function and the fractal dimension to validate his model. A major drawback of these methods is that they do not account for spatial relations.

Spatial similarity evaluates the spatial correspondence between the simulated results and reference land-use maps. Li and Yeh (2001) used a conversion matrix to estimate the accuracy of their simulation. Ménard and Marceau (2007) compiled landscape metrics and spatial statistics to evaluate the ability of their CA model to characterize trends in deforestation. White (2006) used fuzzy polygon matching that is based on the areal intersection of land-use polygons on two maps.

Another method for assessing spatial similarity is the Kappa simulation coefficient, $K_{\text{Simulation}}$. $K_{\text{Simulation}}$ is an index recently proposed specifically adapted for measuring the agreement between categorical maps to overcome the limitations of the standard Kappa coefficient (van Vliet et al. 2011). Hasbani et al. (2011) applied Kappa Simulation statistics to evaluate the quality of their CA model results and demonstrated that Kappa Simulation is very useful to capture differences in the simulation results. $K_{\text{Simulation}}$ expresses the percentage of agreement between two maps by considering the information contained in the initial land-use map and the

proportion of cells that do not change state during the simulation. $K_{\text{Simulation}}$ has two components: $K_{\text{Transition}}$ and K_{Transloc} . $K_{\text{Transition}}$ measures the quantitative similarity between two maps, while K_{Transloc} measures the similarity of the spatial allocation of categories between the maps. $K_{\text{Simulation}}$ has values ranging from -1 to 1, with 1 representing perfect agreement, 0 representing the agreement that can be expected by a random distribution of a certain transition, and values below 0 representing the agreement below what can be estimated by a random distribution of the transition. $K_{\text{Transition}}$ values range from 0 to 1, with 0 indicating that no transitions appear in the simulated map, nor in the historical data, and 1 indicating that the simulated number of transitions is the same as the one in the historical data. K_{Transloc} values range from -1 to 1, where 1 indicates that the allocation is as high as the one in the historical data, 0 represents that the agreement is the same as can be expected in a random allocation, and values below 0 indicate that the allocation of the transitions is worse than what can be expected in a random allocation. Since $K_{\text{Simulation}}$, $K_{\text{Transition}}$, and K_{Transloc} assess the agreement between land-use transitions rather than measure the agreement between two maps as the Kappa index does, it is considered an effective way for evaluating the model results. These three indices were therefore calculated to analyze the simulated results using the Map Comparison Kit developed by the Research Institute for Knowledge System, The Netherlands (RIKS BV 2012).

Using RST for identifying dominant factors provided a solution for the problem of having a large number of factors and land-use classes in a CA model (of which fine-resolution CA is a case). This achieved the goal of tackling the first challenge of CA modeling at fine resolution. RST was further used in a fine-resolution dataset (Chapter 3) that represents a more complex and dynamic landscape involving a considerably larger number of land-use classes, land-use

transitions, and driving factors. The second challenge of building a fine-resolution CA involves defining suitable space representation and simulation procedure, which is presented in the next chapter.

Chapter Three: Development of a Patch-based CA to Simulate Land-use Changes at Fine Resolution

This chapter describes the methodology for developing a patch-based CA that was designed to enable the simulation of land-use changes at fine spatial resolution by integrating the patch concept into the model. In this CA model, a transition probability map is calculated at each cell location for each land-use transition using a weight of evidence method; then, land-use changes are simulated by employing a patch-based procedure based on the probability maps. A traditional cell-based model was also developed for comparison.

3.1 Study area and dataset

3.1.1 Study area

To develop a fine-resolution CA model, this study focuses on a small area in the eastern Elbow River watershed located at the fringe of the fast growing city of Calgary (Figure 3-1). The area covers approximately 250 km² and is within the Trans-Canada Banff-Calgary corridor. According to various municipal development plans, this area is under considerable pressure for land development and is seen as a specific growth-node area due to its proximity to the city of Calgary (East Springbank Area Structure Plan, 2010; IDP, 2011). The area contains three highways, namely, Highway 22 to the west, the Trans-Canada Highway 1 to the north, and Highway 8 to the south. It is characterized by agriculture land (63%) including rangeland and cropland, forest (12%), and built-up areas (25%). The built-up class includes low-density country residential (density is usually lower than 0.5 units per acre), high-density urban residential (density is usually greater than 6 units per acre), commercial, and recreational areas (i.e., golf

courses). With its convenient proximity to the city of Calgary, this area has undergone dramatic land-use changes over the last decade and is under considerable pressure for further development, particularly along the three highways (IDP, 2011).

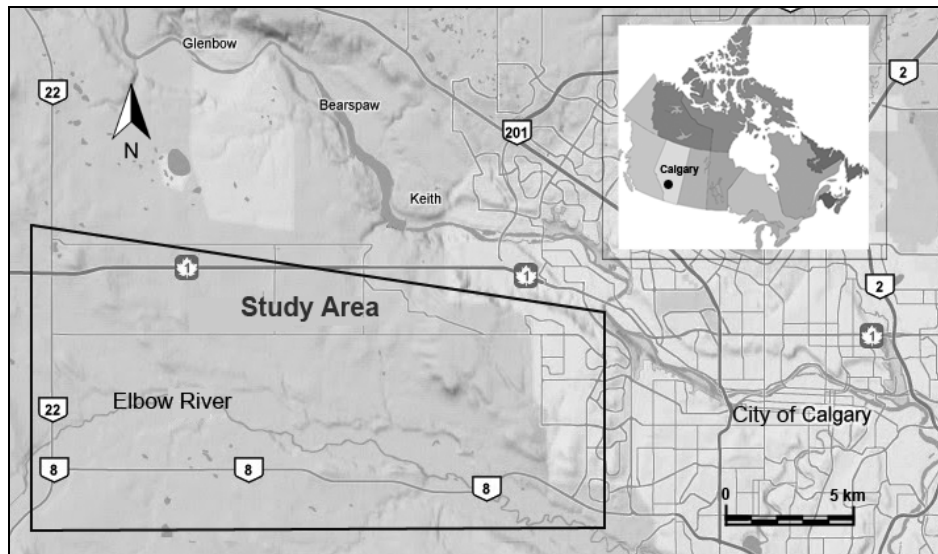


Figure 3-1 Study area (shown in the black box)

3.1.2 Data sources

The main datasets used in this study include four land-use maps generated from SPOT-5 remote sensing images, which provides the highest resolution images available for the study area with historical consistency. SPOT-5 was launched in 2002 and offers high spatial resolution of 2.5 m to 5 m in panchromatic (PAN) mode and 10 m in multispectral (MS) mode. SPOT-5 images can provide fine enough information to discriminate almost all the land-use/land-cover types. Recently, several researches (Kressler et al. 2003, Haeberlin et al. 2004, Duro et al. 2012) have used SPOT-5 data to generate land-use maps and have been successful in discriminating detailed land-use classes, as well as their geometry. SPOT-5 multi-spectral (MS) and panchromatic

(PAN) images at the respective spatial resolution of 10 m and 5 m were acquired for the years of 2003, 2006, 2008, and 2011. To guide the supervised classification of the remote sensing images, field work was conducted in June 2008, during which a total of 255 points distributed over the whole study area were collected and the land-use of each location was recorded.

Additionally, GIS data were obtained from the Calgary Regional Partnership, including a Digital Elevation Model (DEM) at a spatial resolution of 30 m, the city of Calgary's boundary vector file, and the road network in vector format. A slope map was generated from the DEM. The distance to Calgary's boundary was calculated and stored as a raster image at 5 m resolution. Three raster images of distance to three categories of roads, also at 5 m resolution, were created using the *Eucdistance* function in ArcGIS (ESRI, 2012). These datasets were used as external factors for calibrating the CA model.

3.1.3 Land-use map generation

The SPOT-5 remote sensing images were pre-processed and classified to generate the land-use maps, which act as the basis of the CA model calibration and simulation. The pre-processing includes image fusion of the PAN images and the corresponding MS images, as well as georegistration to correct for geometric error in the images. Classification consists of segmentation and an object-based classification procedure.

3.1.3.1 Image fusion and georegistration

Image fusion is a combination of two or more different images to form a new image (Pohl and Van Genderen 1998). One typical application of image fusion is to combine the geometric details

of high-resolution PAN images and the multi-spectral low-resolution MS images to produce a high-resolution MS image. In this project, the SPOT-5 5 m PAN images and 10 m MS images were fused to generate 5 m resolution MS images using the remote sensing data processing software ENVI 4.6 (Exelis Visual Information Solutions 2012). HSV (Hue, Saturation, Value) is a straightforward way to undertake image fusion. During the HSV image fusion process, the low resolution RGB image is converted to HSI (Hue, Saturation, intensity) space. The panchromatic band is then matched and substituted for the intensity band, and finally the HSI image is converted back to RGB space. HSV was selected to process the SPOT-5 images in this study. This method generates histogram information similar to the one in the original images.

Images from different sensors or acquisition dated have different distortions due to changes in location, view angle, refraction, and sensor differences when the images were acquired. After the data fusion, the images were geo-registered using 150 ground control points selected from a road network dataset acquired from the Calgary Regional Partnership. The geo-registration error was within 0.5 pixel for each image.

3.1.3.2 Image segmentation and classification

With the increasing resolution, increasing details of land use can be identified from the remote sensing images and it is very likely that the neighbouring pixels belong to the same land-cover class as the pixel under consideration (Blaschke and Strobl 2001). Therefore, classification of high resolution remote sensing images should reflect the homogeneous groups of pixels that correspond to the real-world objects. This is especially true for manmade objects such as residential areas, roads, etc. For example, in the case of residential areas composed of individual

buildings, the ground resolution of the fused SPOT data is smaller than their average object sizes. Object classifiers can utilize contextual information as well as spectral information and delineate image objects with relatively similar properties corresponding to real houses. Therefore, an object-based classification method would be a better way to deal with high resolution remote sensing images (Duro et al. 2012).

In this study, object-based classification was adopted to produce the historical land-use maps using *eCognition* software, which contains modules specially designed for handling objects (Trimble 2012). The classification consisted of three main steps: 1) image segmentation, 2) image classification, and 3) post classification. Three different levels of segmentation were applied to identify three levels of objects. At level 1, the average segment size was 10 pixels (250 m^2), which was used to classify buildings and three different levels of roads. At level 2, the average segment size was 20 pixels (500 m^2) for the classification of forest and water areas. At level 3, the average segment size was 35 pixels (875 m^2) for the classification of golf courses, agriculture areas, green areas, and areas under development. The classification procedure resulted in ten land-use/cover classes, namely buildings, roads, forest, water, golf course, agriculture areas, green areas, and under development. A post classification was applied to sub-classify different built-up types (residential, commercial/institutional, airport, industrial) based on the building class acquired from field survey, to divide roads into three different levels (road level 1, road level 2, road level 3) based on the road network dataset, as well as to limit the influence of geo-registration errors and temporal inconsistencies. The class 'residential' was sub-classified into two categories: Country Residential (where the density is above 2 acres per unit) and Urban Residential (where the density is around 0.17 acre per unit). Country Residential is an

urban land-use type that refers to the low-density residential typically found at the ‘rural-urban fringe’ where urban-employed individuals reside (Whitehead 1968). The final classification resulted in 14 land-use classes that were further distinguished into two categories: dynamic classes and static classes (Table 3-1). Dynamic classes include Country Residential, Urban Residential, Forest, Agriculture, and Under Development; these classes will change during the simulation. The remaining classes are static, which means that they rarely change over time, but have high influence on the dynamic classes. The classification accuracy was assessed through field validation using 255 points distributed over the study area. The overall classification accuracies measured by the kappa coefficient for the land-use maps of the years 2003, 2006, 2008 and 2011 are 0.85, 0.86, 0.86, and 0.87 respectively. Most of the error exists between Agriculture and Forest while the urban land-use classes (i.e., Country Residential, Urban Residential, and Under Development) have higher accuracy.

Table 3-1 Land-use classes identified in the generated land-use maps

	Number	Class
Static Land-use Class	1	Commercial/Institutional
	2	Industrial
	3	Airport
	6	Road Level 1
	7	Road Level 2
	8	Road Level 3
	9	Golf Course
	11	Water
	13	Green Area
Dynamic Land-use Class	4	Urban Residential (UR)
	5	Country Residential (CR)
	10	Agriculture (A)
	12	Forest (F)
	14	Under Development (U)

The land-use map of the year 2011 is shown as an example in Figure 3-2. The historical land-use maps of 2003, 2006 and 2008 were used for model calibration, while the land-use map 2011 was used for evaluating the performance of the model.

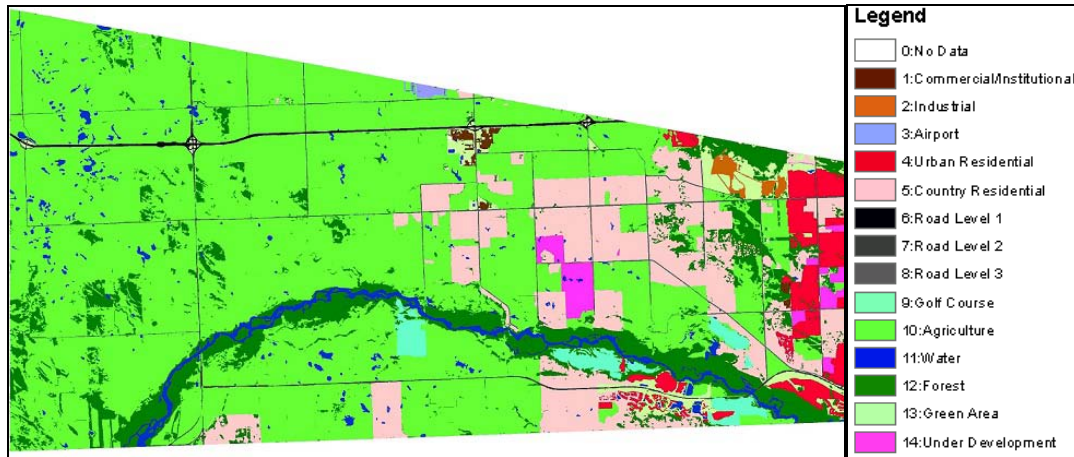


Figure 3-2 Historical land-use map of the year 2011

3.2 Framework of the patch-based CA model

The methodological framework for the development of the patch-based CA model is shown in Figure 3-3 and involves two main components: calibration and simulation. During the calibration process, the dominant factors driving the land-use changes were first selected using a data mining technique, Rough Set Theory (RST). Then the Weight of Evidence (WOE) method was employed for calculating the transition probability at each cell location.

The simulation procedure includes three major parts. First, at each time step, a transition probability map for each type of land-use transition was generated using the WOE approach. Five types of land-use transitions were considered in this study based on observations made from

the historical land-use maps: (i) from Country Residential (CR) to Under Development (U), (ii) from Agriculture (A) to Under Development (U), (iii) from Forest (F) to Under Development (U), (iv) from Under Development (U) to Country Residential (CR), and (v) from Under Development (U) to Urban Residential (UR). The change from Country Residential to Under Development is possible according to the East Springbank Area Structure Plan (2010). This type of change typically occurred through infill development, which intends to create fully serviced Urban Residential at higher density than existing Country Residential, where future land use intensification to urban density levels is desired. Second, based on the observations revealed in the historical data, three types of transitions to the land-use class Under Development happen simultaneously, namely Country Residential, Forest, and Agriculture. Therefore, a unique probability map was created by combining the transition probabilities of these three types of land-use transitions. Third, a patch-based simulation procedure was developed for generating land-use patches based on the transition probability maps. At the end of each time step, a new land-use map was created and was then used as the initial map for the next time step of the simulation. To evaluate the performance of the patch-based CA model, a traditional cell-based CA model was also constructed for comparison.

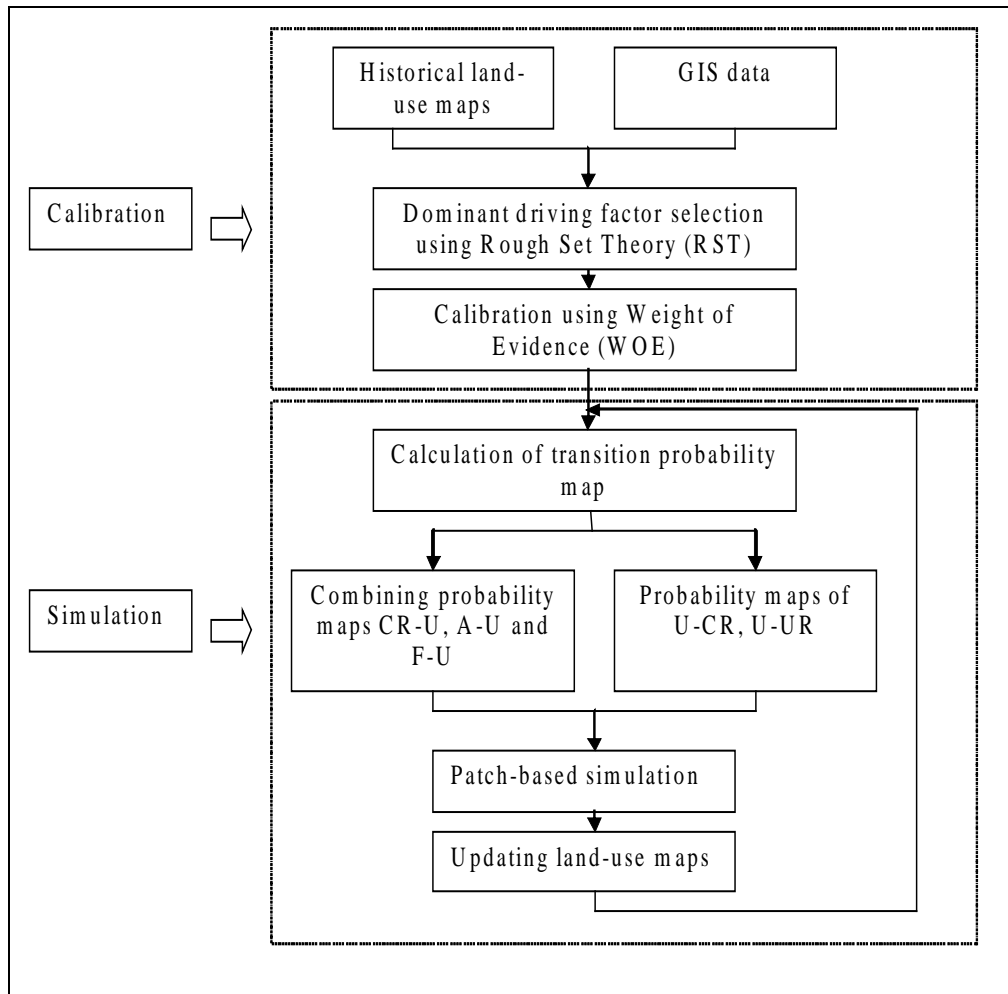


Figure 3-3 Flowchart of the patch-based CA model

3.2.1 Calibration of the patch-based CA model

3.2.1.1 Identification of driving factors

For identifying the significant factors driving land-use changes in the study area, two original sets of factors were considered: internal factors (i.e., neighborhood influence defined by the cell states, extracted from the historical land-use maps) and external factors (i.e., extracted from external data sources, such as slope). To determine the internal factors, a definition of the neighborhood is required. Several neighborhoods are used in CA models, including local and

extended neighborhoods in combination with a variable geometry (e.g., square, circle, and ring). In our project, we defined the neighborhood as composed of three concentric rings following the approach proposed by Hasbani et al. (2011). This neighborhood takes into consideration local and non-local information, while reducing the bias from distant cells by separating the neighborhood into concentric rings of various distances.

In our fine resolution datasets, the cell size is typically much smaller than the real world entities to be simulated. Each entity is represented as a cluster of cells (i.e., a patch) having the same land-use class in the land-use maps. Through an analysis of the historical land-use maps, we calculated the *average radius* (AR) within patches for each land-use class and the average distance between different land-use patches. Then, the three neighborhood ring radii were defined as: (i) the minimum AR, corresponding to 23 cells in this study, (ii) the maximum AR, corresponding to 77 cells, and (iii) the average distance between patches, corresponding to 165 cells. The minimum AR and maximum AR allowed the model to take into account the internal characteristics of the entities, while the average distance between patches enabled the model to consider the relationship between patches. Calculated in meters, the sizes of the radii for the three neighborhoods used in this study are 115 m, 385 m, and 825 m, respectively. It should be noted that the class of Agriculture was not considered for determining the ring radius, as it is composed of large patches, which proved ineffective for determining the major relationship among the land-use classes from our preliminary test. Finally, the number of cells of each land-use class within each neighborhood was extracted, for a total number of 42 factors (i.e., 14 classes by 3 neighborhood radii). The external factors considered in this research include slope, distance to road level 1 (Highway 1), distance to road level 2 (Highway 8 and Highway 22),

distance to road level 3 (other main roads in the study area), and distance to Calgary’s boundary. Such accessibility factors have been proven to be important in many previous land-use CA simulations (Dietzel and Clarke 2007, Liu et al. 2010).

The aforementioned factors are prone to redundancy at the cell scale used, and some of them may not be necessary for developing robust CA models. Rough Set Theory (RST) is a data mining technique (Pawlak 1982) that facilitates the factor selection for the calibration of CA models (Wang et al. 2011). This technique derives an optimal set of factors from an original set while minimizing the redundancy and retaining the original factors. Following the methodology developed in Chapter 2, RST was applied to select the dominant factors for the calibration of the CA model. The selected factors are shown in Table 3-2.

Table 3-2 Selected factors using Rough Set Theory

Initial Land-use Class	RST Selected Factors
CR	Dist2cityBdry; Dist2roadL1; Dist2roadL2; Dist2roadL3; NH0#4; NH0#5; NH0#10; NH0#12; NH1#4; NH1#5; NH1#8; NH1#10; NH1#11; NH2#4; NH2#7; NH2#8
A	Dist2cityBdry; Dist2roadL1; Dist2roadL2; Dist2roadL3; Slope; NH1#4; NH1#5; NH1#8; NH1#10; NH1#12; NH2#11; NH2#12; NH2#14;
F	Dist2cityBdry; Dist2roadL1; Dist2roadL2; Dist2roadL3; Slope; NH0#5; NH0#12; NH1#8; NH1#10; NH1#11; NH1#12; NH1#14;
U	Dist2cityBdry; Dist2roadL1; Dist2roadL2; Dist2roadL3; NH0#4; NH0#5; NH0#14; NH1#4; NH1#5; NH1#8; NH1#10; NH1#14; NH2#4; NH2#5; NH2#8; NH2#10; NH2#14;

3.2.1.2 Calibration using WOE

Adequately defining the transition rules is a critical step in CA modeling and numerous techniques have been proposed, including statistical methods, such as logistic regression (Wu 2002), transition probability functions (Jenerette and Wu 2001), and computational intelligence methods such as artificial neural network (Almeida et al. 2008), genetic algorithm (Shan et al. 2008), and support vector machines (Yang et al. 2008). Among them, the transition probability function has been widely used due to its simplicity, clarity of the parameters, and easy interpretation of the results. The general format of a transition probability function is defined as:

$$S_{xy}^{t+1} = f(P'_{xy}) \quad (3-1)$$

where, S_{xy}^{t+1} is the land-use state in the location (x, y) at time step $t+1$, and P'_{xy} is the transition probability of state S in location (x, y) at time step t . According to transition probability rules, the change of a cell's state is based on the probability value at that location. This value was calculated for each dynamic land-use class at each time step based on the RST selected factors. The higher the value is, the higher the chances are that the transition will occur. The cells were ranked by their probability values for each type of land-use change. The simulation begins by locating the cells with the highest values and proceeds downwards until a sufficient number that meets the requirement from historical data of cells' state changes have been achieved. Transition probability values are dynamically updated at each time step.

The transition probability function is usually the combination of a series of weighted factors (Santé et al. 2010). In this study, WOE was chosen to calculate the transition probability function. Based on Bayesian theory, WOE can be used to combine information for a set of

discretized factors from known locations, and calculate the transition probability of the unknown locations (Bonham-Carter et al. 1989). This method has demonstrated to be robust even with small sample sizes over a large area, it does not rely on the assumption that the input data are normally distributed, and the implementation and interpretation of the weight values is very intuitive (Dickson et al. 2006). WOE has been widely used in medical diagnoses (Spiegelhalter 1986), spatial analysis and prediction of mineral deposits (Raines and Mihalasky 2002), and plant migration prediction (Lyford et al. 2003). Based on their urban growth simulation of Sao Paulo State, in Brazil, Almeida et al. (2003) demonstrated that WOE has the capacity to accommodate nonlinear characteristics of a complex urban system. Thapa and Murayama (2011) adopted WOE in their CA model to simulate urban land-use changes in Kathmandu metropolitan region, Nepal, and demonstrated that the WOE approach could effectively handle categorical maps.

WOE applied for calculating transition probabilities follows the methodological procedure described by Bonham-Carter et al. (1989). The posterior probability of a transition (i.e., $s \rightarrow k$) is calculated using:

$$P\{s \rightarrow k | B \cap C \cap D \dots \cap N\} = \frac{e^{\sum W_N^+}}{1 + e^{\sum W_N^+}} \quad (3-2)$$

where, B , C , D , and N are the values of the factors,

W_N^+ represents the weight for each factor.

The weights of the factors are assigned based on the prior probability of factors at the known locations.

$$W^+ = Ln\left(\frac{P(E_i|D)}{P(E_i|\bar{D})}\right) \quad (3-3)$$

$$P(D) = \frac{N(D)}{N(T)} \quad (3-4)$$

where $N(D)$ is the number of the transitions that occurred,

$N(T)$ is the total number of the cells, and

$P(D)$ is the prior probability of the transition.

Since WOE only applies to categorical data, an important step is to break the values of each factor into a series of ranges (categories), following the method described by Agterberg and Bonham-Carter (1990). First, the relationship curve between the values of the factor and the number of transitions that occurred in the historical data was constructed. Then, breaking points on the curve were determined by creating a best-fitting curve using a series of straight-line segments employing a line-generalizing algorithm. A weight was estimated for each range of each factor using the following function:

$$W^+ = \ln\left(\frac{y_{n=k} - y_{n=k-1}}{A_{n=k} - A_{n=k-1}}\right) \quad (3-5)$$

$$y_n = A_n * \exp(W^+) \quad (3-6)$$

where n is the number of ranges for each factor, k represents the breakpoints defined for the ranges, and A_n is the number of cells for each range.

The WOE method estimates the probability of change at each location based on prior information that comes from the historical data. In this study, the evidence consists of a set of factors and the hypothesis is “the location is favorable for occurrence of land-use change”. Weights are estimated from the measured association between known occurrence of land-use changes and the values of the factors, which will then be used for calculating future land-use change probability at each location, producing a transition probability map.

At each time step, a transition probability map for each type of predefined land-use change was calculated at each cell location using the WOE function with the initial land-use map and the values of the selected factors as the input. Here, the transition probability maps were only calculated for the predefined land-use transitions between the dynamic land-use classes. Static land-use classes act as factors influencing the transitions between the dynamic land-use classes. Figure 3-4 provides an example of such a map for the land-use transition from Agriculture to Under Development. It can be observed that the probability is higher in areas that are adjacent to existing development.



Figure 3-4 Probability map for the transition from Agriculture to Under Development (brighter tone showing higher probability values)

3.2.2 Simulation module of the patch-based CA model

3.2.2.1 Transition probability map combination

Compared to what can be observed at coarse spatial resolutions, most of the geographic entities defined on fine resolution land-use maps are composed of multiple cells. The objective of land-use simulation at fine resolutions is to locate patches representing meaningful entities rather than individual cells (Chen et al. 2012). Also, at fine resolutions, different land-use transitions often happen simultaneously (Wu 2002). For instance, the transition from Country Residential to Under Development rarely occurs alone, as the Country Residential area is often closely surrounded by Forest and/or Agriculture. An examination of the historical data also shows that

the three types of land-use transitions (i.e., CR to U, A to U, and F to U) tend to occur simultaneously. An accurate simulation requires considering the relationship between these three types of land-use changes.

In this project, we generated a unique probability map by combining the transition probability maps for these three types of transitions to Under Development (i.e., CR to U, A to U, and F to U). Equal weights for combining the three types of transitions were used. The simulation of the land-use patch Under Development was based on the combined probability map.

3.2.2.2 Patch-based simulation

In this study, the traditional individual cell-based simulation procedure was replaced by a patch-based simulation process to identify the appropriate patches that might change states. The simulation procedure was performed using five main steps. First, four parameters indicating the patch information about the three final land-use classes (U, CR, and UR) were extracted from the historical land-use maps including mean patch size (MPS), minimum patch size (MiPS), number of changed cells (NCC), and number of patches (NOP). Second, potential patches were formed based on the derived patch information through a seed filling strategy. For the transition to Under Development, the seeds of the patches corresponding to the cells having the highest values in the combined transition probability map were selected. For the transitions from Under Development to Country Residential and to Urban Residential, the seeds of the patches were selected from the original transition probability maps generated by the WOE method. Then the cells with a lower level of probability values around the seed cells were selected and merged with seed cells to form patches. This process was repeated until the number of the potential cells for the patch

reaches a certain number. In this project, it is 2.5 times NCC. This number was calculated based on the patch information extracted from the historical land-use maps. Third, small patches were removed based on the MiPS found in the historical maps. Fourth, a mean transition probability value was calculated for each patch. Finally, the mean probability values for the patches were ranked; the patches with higher mean probability values were selected and their states were updated. The number of patches changed is equal to NOP found in the historical data. This procedure was repeated for each type of land-use transition resulting in a new land-use map.

3.2.3 Development of a cell-based CA model

To evaluate the performance of our patch-based model, a traditional cell-based CA was developed at 5 m resolution for comparison. This CA was also based on the transition probability maps derived from the WOE method. For each type of land-use transition, the values of the transition probability maps at each cell location were ranked. The set of cells with the higher probability values were selected to change their cell states accordingly. The transition probability threshold was determined by calculating the number of cells that had changed for each type of land-use transition using the historical land-use maps. Since the two models use the same set of transition maps, their results are comparable.

3.2.4 Evaluation of the results

Model validation was done by comparing the results of the patch-based and the cell-based CA models with the historical reference maps using visual comparison, landscape metrics, and $K_{\text{simulation}}$ indices.

As demonstrated by Hagen-Zanker and Lajoie (2008), a common approach to assess the quality of the results is to compare the simulated map with an independent reference map using landscape metrics, which is an attractive way for measuring land-use patterns (Peng et al. 2010, Aguilera et al. 2011). Since our objective is to evaluate the performance of a land-use patch-based CA model, three landscape metrics were used, namely the number of patches (NOP), the mean patch size (MPS), and the patch size coefficient of variation (PSCOV) defined as the ratio of the standard deviation to the mean patch size. This information was analyzed using Fragstats 3.4, a tool designed to compute a variety of landscape metrics (McGarigal et al. 1995). The $K_{\text{simulation}}$ indices were calculated using the Map Comparison Kit (RIKS BV 2012).

Chapter Four: Scenario Evaluation with the Patch-based CA Model

Scenario evaluation is widely used in land-use planning as a critical tool for foreseeing the potential future states of regions of interest (Jantz et al. 2010, Francis and Hamm 2011). In this chapter, the previously developed patch-based CA model was applied combined with information extracted from the local municipal development plans to simulate three land development scenarios in the eastern part of the Elbow River watershed.

4.1 Context of study

In 1902, oil was discovered in Alberta and it has become an important industry since 1947. When the oil prices increased in 1973, the economy of the city of Calgary in the south of the province grew quickly. The city's population also increased from 403,000 in 1971 to 675,000 in 1989 (Wikipedia 2012). In each decade since 1961, the population has increased by approximately 35% to the current population of 1,043,000 inhabitants and the city's municipal lands have expanded by 14% to the current area of 850 km² (Applied History Research Group 1997-2001, City of Calgary 2008, Statistics Canada 2009). Currently, Calgary is the fourth-largest city in Canada and one of the most attractive labor markets in the country, with economic activities centred on the petroleum industry, tourism, and agriculture (Statistics Canada 2009). It is predicted that Calgary's population will reach about 1.6 million by 2037 and two million inhabitants by 2100 (City of Calgary 2009). Most of this growth will occur primarily due to immigration (Cooper 2006).

As a consequence of such rapid population growth, the city and its surroundings are experiencing intensive land-use changes, a phenomenon that considerably influences the adjacent environmentally sensitive areas. In particular, the western fringe of Calgary, located in the eastern portion of the Elbow River watershed and only 80 km from Banff National Park, is under intensive pressure for land development due to its convenient transportation system and proximity to the city of Calgary. Since 1989, over 1,000 lots have been created for residential purposes, with 65.8% of them being two to four acres per unit in size. Such low density and non-contiguous development also occur farther west in the watershed, causing loss of productive agricultural lands, forest cover, surface water bodies, and increasing levels of water pollution (Central Springbank Area Structure Plan 2001). One critical concern is the sustainability of water resources in the watershed that supplies about 40% of the drinking water for the city of Calgary (Parks Foundation Calgary 2008), supplies irrigation water for agriculture, and fulfills water demands for industry and recreation.

The eastern Elbow River watershed covers about 250 km² with about 14% located in the city of Calgary and 86% located in the Municipal District of Rocky View No. 44. It offers a unique and diverse mixture of natural features, rural elements and urban components, with the main coverage being agriculture, forest, low-density country residential (density is usually lower than 0.5 unit per acre), high-density urban residential (density is usually greater than 6 units per acre), commercial, and recreational area. Historically, the major role of this area is a productive and vibrant agricultural area dominated by rangelands and croplands. This area has also been identified as a potential growth region according to the Municipal Development Plan (1998), Central Springbank Area Structure Plan (2008), and Rocky View County/City of Calgary

Intermunicipal Development Plan (2011). Figure 4-1 shows the two main growth corridors identified from the Rocky View County/City of Calgary Intermunicipal Development Plan (2011). Land-use changes have dramatically affected the environment in the study area, especially the agriculture and water resources. Investigation has shown that both the water quantity and water quality have decreased, so has the ecological diversity (Sosiak and Dixon 2004, Wijesekara et al. 2012, ERWP 2012). Therefore, it is vital for scientists and planners to understand current and future land use in order to balance land development with sustainable growth, and limit unintended environmental problems.

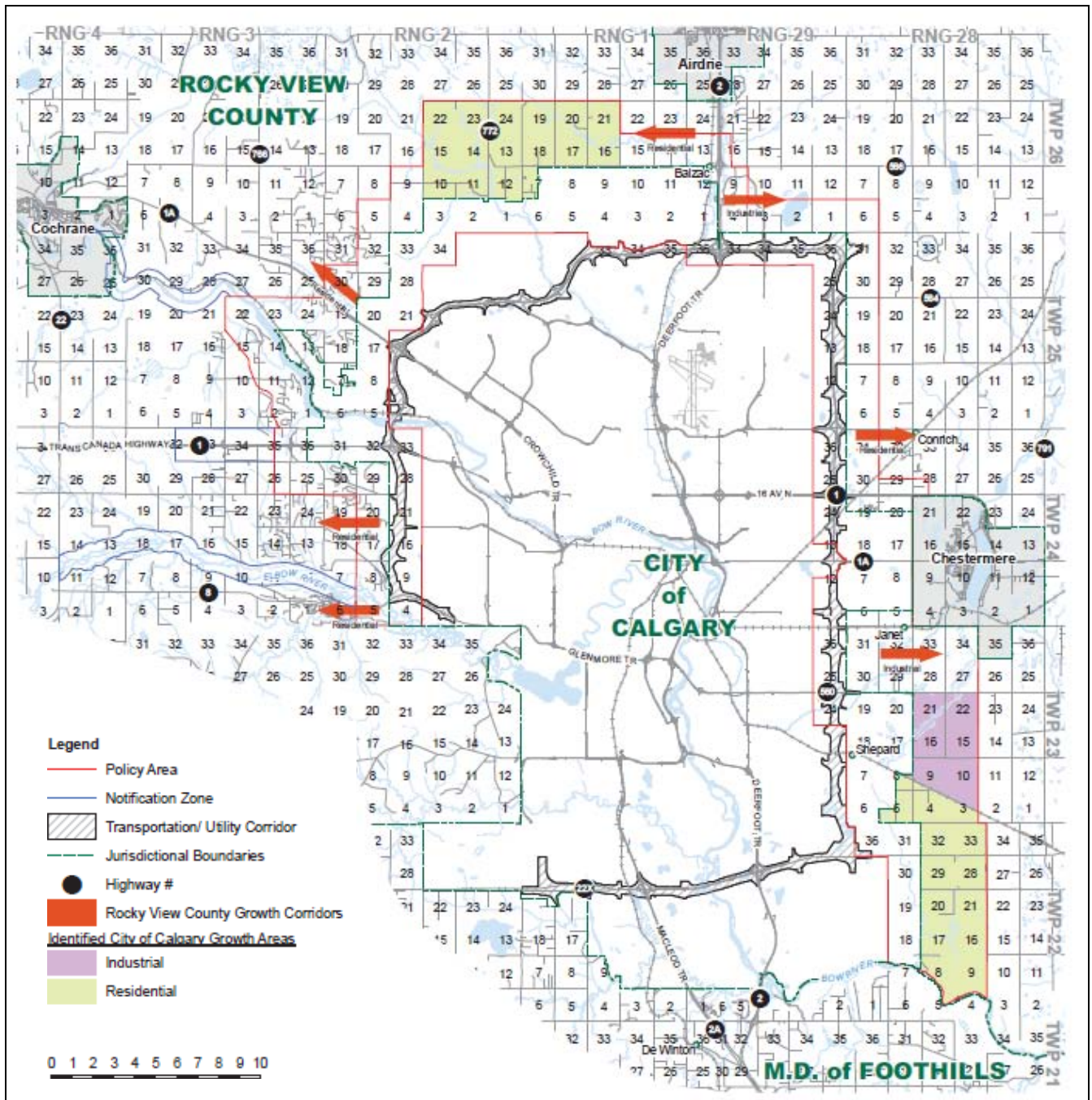


Figure 4-1 Growth corridor for the eastern Elbow River watershed (IDP, 2011)

Scenario evaluation is a process of analyzing possible future events by considering alternative possible outcomes (Wen et al. 2005). It is a tool for balancing land-use changes with sustainable

growth, examining the emerging spatial patterns of the scenarios, and facilitating decision making. Many researchers argue that scenario evaluation should be a key focus for land-use modelling in order to test different land-use planning practices (Torrens and O'Sullivan 2001, Yang and Lo 2003, Thapa and Murayama 2012). Numerous studies have been conducted regarding scenario evaluation. For example, Solecki and Oliveri (2004) defined land-use scenarios using Intergovernmental Panel on Climate Change (IPCC) regional greenhouse gas (GHG) emissions scenarios in the New York metropolitan region, USA. Conway and Lathrop (2005) modeled the ecological consequences of four scenarios in New Jersey, USA. He et al. (2006) modeled the growth of the greater Beijing area using different urban land demand scenarios under different restrictions of water shortage. Shen et al. (2009) tested two different scenarios of population densities (high and low) to simulate future land-use systems in Hong Kong. Zhang et al. (2011) investigated three different scenarios (baseline, service oriented center, and manufacturing dominant center) using Markov chain analysis and CA to understand the urban dynamics of Shanghai, China. Norman et al. (2012) used a popular CA model, SLEUTH, to simulate three different scenarios in a binational dryland watershed. Scenario evaluation can aid land-use management, develop suitable plans before irreversible land-use conversions take place, and guide land-use planning for sustainable development (Zhang et al. 2011).

In recent years, there has been an increasing interest for developing sustainable land-use systems. A sustainable land-use system is defined as a system 'that improves the long-term health of human and ecological systems' (Wheeler 2000). The main concern of land-use sustainability is to better preserve open space and sensitive ecosystems such as agricultural lands (Robinson

2011). Four operational principles for sustainable land-use systems include: 1) not to convert too much agriculture land during the early development stages; 2) to determine the amount of land consumption based on anticipated population growth; 3) to develop the sites which are less important for food production; and 4) to maintain compact development patterns (Plan It Calgary 2007). In this chapter, the above criteria were combined with the patch-based CA model developed in Chapter 3 to simulate land development under three alternative scenarios based on current trends, and management goals of environmental protection and projected population growth in the eastern Elbow River watershed. Section 4.2 describes the additional data required for the scenario testing. Section 4.3 presents the methodology for defining various scenarios. while section 4.4 presents the simulation procedure.

4.2 Data

In addition to the datasets described in section 3.1.2, two datasets were required for this study. The first is *land capacity for agriculture*, which was acquired from the Canada Land Inventory (2008) at the original resolution of 6 m (Figure 4-2). It was resampled to 5 m to be compatible with the historical land-use maps described in Chapter 3. Different colors in the land capacity for agriculture map represent different classes of agriculture suitability, where a small number represents higher suitability and a large number represents lower suitability. For example, the class 2 in the map represents a moderate limitation for agriculture, class 3 represents moderately severe limitations for agriculture, and class 7 represents little capacity for agriculture.

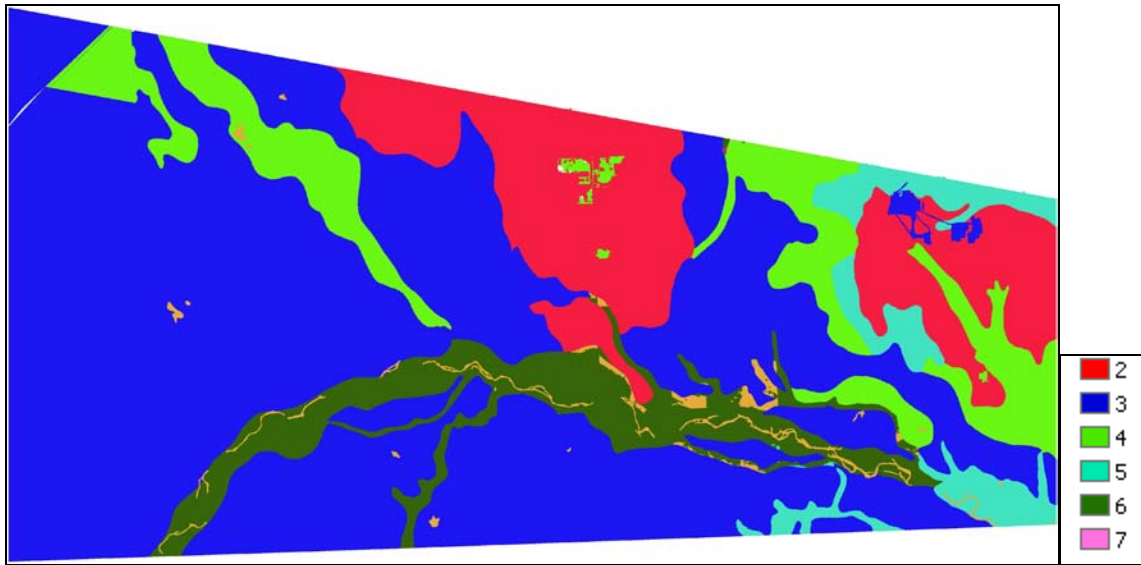


Figure 4-2 Land capacity for agriculture (Canada Land Inventory 2008)

The second dataset is the flood hazard map (Figure 4-3) of the Elbow River acquired from the Government of Alberta (2012). This map provides general information of the flood hazard around the Elbow River as a guide to constrain development from the hazardous area. The flood hazard area is typically divided into floodway and flood fringe zones. Floodway is the portion of the flood hazard area where flows are deepest, fastest, and most destructive. The floodway typically includes the main channel of a stream and a portion of the adjacent overbank area. New development is discouraged in the floodway and may not be permitted in some communities. Flood Fringe is the portion of the flood hazard area outside of the floodway. Water in the flood fringe is generally shallower and flows more slowly than in the floodway. New development in the flood fringe should be limited.

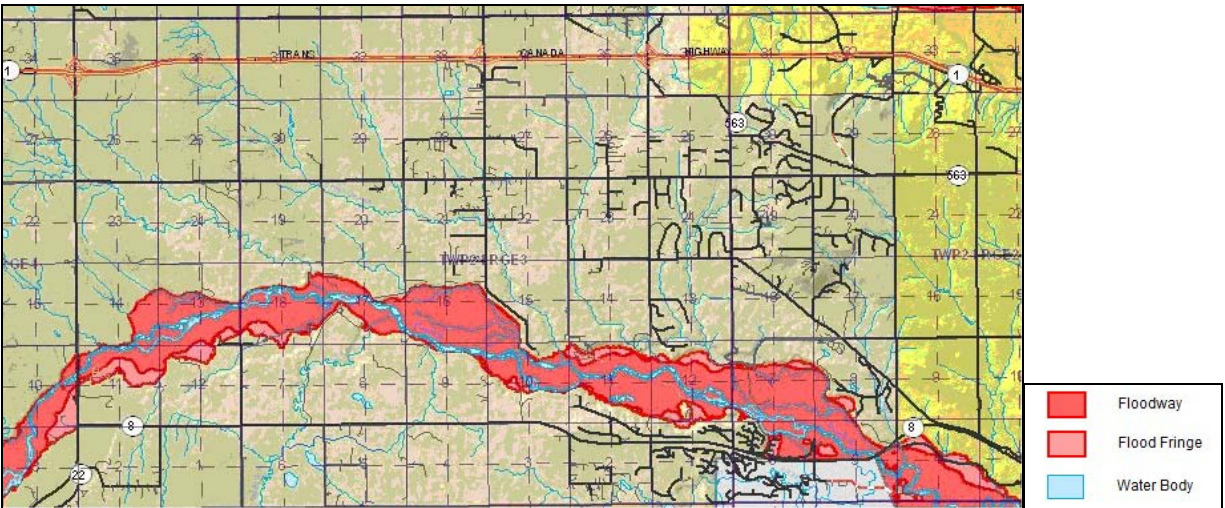


Figure 4-3 Flood hazard map of the Elbow River (Government of Alberta 2012)

4.3 Scenarios definition

The previously defined patch-based CA model (Chapter 3) for the eastern Elbow River watershed was applied for testing various scenarios by adjusting model parameters. Six parameters (Table 4-1) controlling the land-use changes need to be specified for defining a scenario. These parameters are categorized into two groups: the number of changed cells that constrain the total number of *cells* to be changed for each type of land-use transition, and the number of changed *patches* that define the total number of patches to be changed for each type of land-use transition.

Table 4-1 Parameters that need to be specified for defining a scenario

Number	Parameter
1	Number of changed cells for land-use changes to Under Development
2	Number of changed cells for land-use changes from Under Development to Urban Residential
3	Number of changed cells for land-use changes from Under Development to Country Residential,
4	Number of changed patches for land-use changes to Under Development
5	Number of changed patches for land-use changes from Under Development to Urban Residential,
6	Number of changed patches for land-use changes from Under Development to Country Residential.

According to the information extracted from the local municipal development plans, three different scenarios were designed: the *business-as-usual scenario*, the *protective growth scenario*, and the *smart growth scenario*, each of which is linked with different growth conditions for future land-use development.

4.3.1 Business-as-usual scenario

The *business-as-usual scenario* assumes that the future land-use change rate remains unchanged when environmental and developmental conditions are similar to the ones observed from the historical data. This scenario was used to give insight into the spatial consequence of land-use changes under the same initial conditions as those used for the past to present simulation. That is, the same initial conditions as those used to simulate from the past (2003) to present (2011) were used for this scenario. The input parameters for this scenario are listed in Table 4-1.

Table 4-2 Parameters for the *business-as-usual scenario*

Change type	Number of changed cells	Minimum patch size	Number of changed patches
Change to Under Development	83513	5841	13
Change from Under Development to Urban Residential	41711	5845	7
Change from Under Development to Country Residential	41471	6042	6

4.3.2 Protective growth scenario

The *protective growth scenario* aims at promoting land development in a more sustainable way by considering watershed and agriculture protection, as well as use less land resource while aiming at accommodating the same amount of people as the one described in the *business-as-usual scenario*. Three aspects were considered when implementing this scenario.

First, land-use changes affect the water quality and quantity inside the watershed, which are important for the operation of the agriculture system, providing a safe drinking water supply, and supporting a healthy and functional ecosystem. According to the Municipal District of Rocky View/City of Calgary Intermunicipal Development Plan (IDP 2011), all developments proposed in proximity to water bodies should be carefully evaluated. Therefore, one important task during long-term sustainable development is to mitigate the negative impact on water quality and water quantity. To protect the water source and prevent the development of areas located in proximity to the river, in this scenario, a buffer zone of 120 to 200 m wide was created from the center line of the river to cover not only the fresh water, but also the flood plain and the forest along the

river, which are of great ecological value. Development that falls within this buffer zone was excluded.

Second, agriculture is a dominant land use within the area and will continue to have a strong presence in the area. Therefore, a conscious effort and dual responsibility to both the present and the future community will be an important factor in maintaining the viability of agriculture in the study area (Municipal District of Rocky View No.44 1998). The conversion to non-agricultural uses must be integrated with and respectful of agricultural operations. One goal of this scenario is to reduce the development pressure on the lands with high viability for the agricultural industry in the Municipal District of Rocky View No. 44. To fulfill this goal, the sites with high agriculture capacities were excluded from development using the land capacity for agriculture map from Canada Land Inventory (2008). Agriculture capacities classified as class 1 (no significant limitations) and class 2 (moderate limitations) were considered as good quality agriculture land and land development of these sites was excluded. Therefore, the loss of good quality agricultural lands can be avoided.

The last important goal is to conserve more non-developed land. To achieve this goal while allowing the same amount of people to be settled in the area, one way is to increase the area of high density development, while decreasing the area of low density development. The Plan It Calgary (2007) workbook also pointed out that to accommodate the large population growth and keep the sustainability of the city of Calgary and the adjacent areas, more efficient land use types, especially high-density residential, will be encouraged. Specifically, by 2036, land use

efficiency will increase by at least 30 percent as measured by increased density. Here, land-use efficiency is defined as the amount of land needed for accommodating a certain amount of people. Calculated based on the information above, the detailed input parameters for the *Protective Growth Scenario* are listed in Table 4-2.

Table 4-3 Parameters for the *protective growth scenario*

Change type	Number of changed cells	Minimum patch size	Number of changed patches
Change to Under Development	55675	5841	9
Change from Under Development to Urban Residential	44116	5845	7
Change from Under Development to Country Residential	11559	6042	2

4.3.3 Smart growth scenario

The *smart growth scenario* simulates *smart growth*, a contemporary urban planning approach that encourages the sustainable development of cities through a compact urban form in order to conserve more land (Plan it Calgary 2007). This scenario takes into consideration the projected population and economic growth in the city of Calgary and surroundings, with the objective of increasing the land-use efficiency of the development by 30% as suggested by Imagine Calgary Plan for Long Range Urban Sustainability (2006). To increase land-use efficiency, a major strategy is to increase the development of high-density residential (i.e., Urban Residential) and decrease the development of low-density residential (i.e., Country Residential), which then reduce the per capita demand for the occupied land.

To calculate the input parameters for this scenario, the population and dwelling statistics for the Rocky View County and the city of Calgary were used (Table 4-3). According to Statistics Canada (2012), the cumulative growth for the city of Calgary from 2006 to 2011 is 10.9% while the cumulative growth rate for Rocky View County is 9.9%. Based on this information, an average population growth rate of 2% per year was applied in the study area, which corresponds to about 2643 people at each time step of the simulation (i.e., 3 years).

Table 4-4 Population and dwelling statistics for the Rocky View County and city of Calgary (Statistics Canada 2012)

Geographic Name	CSD Type	Population			Private Dwellings, 2011	
		2011	2006	Change Rate (%)	Total	Occupied by Usual Residents
Rocky View County	Municipal District	36,461	33,173	9.9	12,810	12,077
Calgary	City	1,096,833	988,812	10.9	445,848	423,417

When dividing the population by the dwellings in 2011, for both the city of Calgary and the Rocky View County, each dwelling has an average of three people. For the class Country Residential, each dwelling covers above 2 acres per unit, while for the Urban Residential, covers about 0.17 acres per unit. In other words, each person in the Urban Residential occupies about 229 m² (corresponding to 9 cells in the CA model) and one person in the Country Residential area occupies about 2698 m² (corresponding to 108 cells). To accommodate a total growth of 2% of the population in the study area, different ratios of Country Residential and Urban Residential area can exist, resulting in different land-use efficiencies.

Calculated from the historical data, the ratio of the current population settling in the Urban Residential and in the Country Residential is about 12:1. This ratio needs to be improved to 37:1 in order to improve the land-use efficiency by 30%. To accommodate the projected population increase at each time step of the simulation, the number of changed cells and the number of changed patches for each type of land-use transition is listed in Table 4-4.

Table 4-5 Parameters for the *smart growth scenario*

Change type	Number of changed cells	Minimum patch size	Number of changed patches
Change to Under Development	30740	5841	6
Change from Under Development to Urban Residential	23180	5845	5
Change from Under Development to Country Residential	7560	6042	1

4.4 Simulation

Simulations of these scenarios were run from 2011 to 2041 with a time step of three years and the historical land-use map of 2011 as the initial map. The type of land-use classes and land-use transitions simulated in all the three scenarios remain the same as the simulation from past to present described in Chapter 3.

Chapter Five: Results and interpretation

This chapter provides the results for the RST factor selection part described in Chapter 2, the patch-based CA model described in Chapter 3, and the scenario evaluation described in Chapter 4.

5.1 Factor selection using RST

5.1.1 Results

The key driving factors identified through RST are shown in Table 5-1. A first observation is that the nine factors selected for Vegetation and the eight factors selected for Forest are not identical. This suggests that different factors are responsible for different types of land-use change. The factor number of Built-up cells in all the three neighborhoods (i.e. NH0, NH1, and NH2) was identified in both the Vegetation and the Forest decision tables, confirming the strong attractiveness of this internal factor. The external factors were the most important factors in both decision tables based on their frequency of identification. Distance to city center, distance to road and distance to river are associated with the transition of Vegetation to Built-up areas, while distance to river and distance to city centre are associated to the transition from Forest to Built-up.

Table 5-1 Selected factors using Rough Set Theory feature selection

Initial cell state	Factors
Vegetation	Distance to City Center; Distance to road; Distance to river; Number of Vegetation cells in NH0; Number of Built-up cells in NH0; Number of Forest cells in NH1; Number of Built-up cells in NH1; Number of Vegetation cells in NH2; Number of Built-up cells in NH2
Forest	Distance to river; Distance to City Center; Number of Built-up cells in NH0; Number of Vegetation cells in NH1; Number of Built-up cells in NH1; Number of Forest cells in NH2; Number of Vegetation cells in NH2; Number of Built-up cells in NH2

The Kappa Simulation indices calculated for each land use are shown in Table 5-2. For the class Built-up, the $K_{Simulation}$ values obtained using the RST selected factors are higher than those using the original 18 factors for all the years, with the greatest improvement (0.135) achieved for the year 1992. For the class Forest, the $K_{Simulation}$ values obtained when using the RST selected factors are also higher than the ones obtained with the original 18 factors except for the year 2001 where the difference in results is small (0.016). Significant improvement is reached for the year 2006 (0.223). For the class Vegetation, the $K_{Simulation}$ values obtained using the RST selected factors are similar to the ones calculated for the original 18 factors with the biggest difference being 0.025 for the year 1992.

These results indicate that the reduced number of factors selected by RST generates an overall land-use distribution that is more similar to the observed distribution in the historical maps than the 18 original factors. Moreover, the computation time required to obtain the simulation results from the RST factors is about one third of the time needed to generate the results using the original set of 18 factors. The reduced performance for the class Vegetation might be explained

by the fact that agriculture and rangeland/parkland have been aggregated within that class, while their respective dynamics might be affected by different driving factors.

The $K_{\text{Transition}}$ values are higher for the simulation results using the RST selected factors for most cases except for the classes Vegetation and Forest for the year 2001 where the difference in results is 0.023 and 0.064 respectively. The highest improvements achieved are for the class Built-up (0.397) in 2001, for the class Vegetation (0.294) in 1992, and for the class Forest (0.475) also in 1992.

However, the K_{Transloc} values are lower for most cases when using the RST selected factors. For the class Built-up, the K_{Transloc} values are similar for the year 1992, but vary moderately for the years 2006 (0.054) and 2001 (0.1), and vary significantly for the year 1996 (0.171). For the class Vegetation, the values are similar for the years 1996, 2001, and 2006, but vary considerably for the year 1992 (0.172). For the class Forest, the values are similar for the years 1996 and 2001, but vary considerably for the year 1992 (0.148) and 2006 (0.157).

Table 5-2 $K_{Simulation}$ statistics for each land-use class for the simulation results obtained with the original 18 factors and the RST selected factors respectively

			1992	1996	2001	2006
Built-up	$K_{Simulation}$	All factors	0.168	0.242	0.155	0.218
		RST factors	0.303	0.271	0.203	0.282
	$K_{Transition}$	All factors	0.399	0.423	0.459	0.578
		RST factors	0.711	0.675	0.856	0.654
	$K_{Transloc}$	All factors	0.421	0.572	0.337	0.377
		RST factors	0.426	0.401	0.237	0.431
Vegetation	$K_{Simulation}$	All factors	0.201	0.267	0.198	0.231
		RST factors	0.226	0.248	0.202	0.238
	$K_{Transition}$	All factors	0.407	0.748	0.556	0.621
		RST factors	0.701	0.772	0.533	0.677
	$K_{Transloc}$	All factors	0.494	0.357	0.357	0.372
		RST factors	0.322	0.321	0.377	0.351
Forest	$K_{Simulation}$	All factors	0.203	0.313	0.362	0.211
		RST factors	0.291	0.329	0.346	0.434
	$K_{Transition}$	All factors	0.431	0.742	0.880	0.621
		RST factors	0.906	0.861	0.806	0.872
	$K_{Transloc}$	All factors	0.469	0.422	0.412	0.340
		RST factors	0.321	0.381	0.429	0.497

These results reveal that the RST selected factors generally provide a better performance at capturing quantity information rather than location information. One explanation is that in the transition rules generated with RST, there are cases where one rule leads to two types of land-use change (e.g. Vegetation to Built-up, and Vegetation to Forest). In the general case, where one transition rule leads to only one type of land-use change (e.g. Vegetation to Built-up), specific locations of the cells that will change states are defined. However, in the particular case where one transition rule leads to two types of land-use change (e.g. Vegetation to Built-up and Vegetation to Forest), the locations for each transition are randomly selected among suitable

cells. In the transition rules generated with the original 18 factors, the chances of having cases where one rule leads to two types of change is smaller than in the transition rules generated using the RST selected factors due to the fact that when more factors are involved in a conditional transition rule, the rule becomes more specific. For example, for the class Forest, there are about 7% of cells in the historical maps corresponding to the transition rules generated with the original 18 factors that lead to changes to both Vegetation and Built-up while there are about 11% of cells corresponding to the transition rules generated with the RST selected factors leading to changes to these same land-use classes. Therefore the chance of having an error of location is slightly greater with the RST factors than with the large set of original factors.

5.1.2 Summary

This study represents a first step in evaluating the potential of a data mining technique, Rough Set Theory (RST), to identify the key driving factors and generate the transition rules required for the calibration of a land-use CA model. It reveals some promising outcomes. First, the factors selected by RST are not identical for each land use, suggesting that different factors are responsible for different land-use changes in the study area. Second, as it could be expected, it reveals that among the identified factors, the external factors (distance to river, distance to city center, and distance to road) and the presence of the Built-up land-use class in the three neighborhoods are key factors driving the transition from Vegetation or Forest to Built-up. The Kappa Simulation indices used to compare the influence of the two groups of factors on the simulation results reveal that the factors selected by RST tend to generate a higher agreement with reference land-use maps than the original group of 18 factors. This is reflected by the overall $K_{\text{Simulation}}$ and $K_{\text{Transition}}$ values for the three land-use classes considered in the study. The

results also indicate that the RST factors are better at capturing quantity information than location information.

One advantage of RST over other calibration methods is that it retains the original factors, which greatly facilitates the interpretation of the results. However, the technique has some drawbacks. First, it can only deal with discretized data. Discretization means introducing the assumption that the original continuous data can strictly fall into a limited number of discrete categories. A sensitivity analysis revealed that using three or four categories for the discretization procedure in this study did not affect significantly the selection of factors, but this aspect requires further investigation. Second, RST is sensitive to potential errors introduced in the original dataset such as attribute and location errors that are almost impossible to avoid when generating historical land-use maps from remote sensing imagery.

5.2 Development of the patch-based CA model

5.2.1 Results

The simulated results and the historical land-use map of the year 2011 are presented in Figure 5-1. A visual analysis shows that the patch-based CA model generates very compact land-use patterns that are similar to the ones observed in the historical map, while the overall patterns generated by the cell-based CA model are more fragmented, as expected. Specifically, the cell-based CA model produces a large number of unrealistic sparsely distributed patches/cells of Under Development in the east part of the study area, which did not appear on the historical land-use maps. This illustrates that the patch-based simulation procedure, which identifies the potential changes based on the mean probability values of patches, rather than individual cells, is

an effective solution in CA modeling for generating compact land-use patterns at fine spatial resolution.

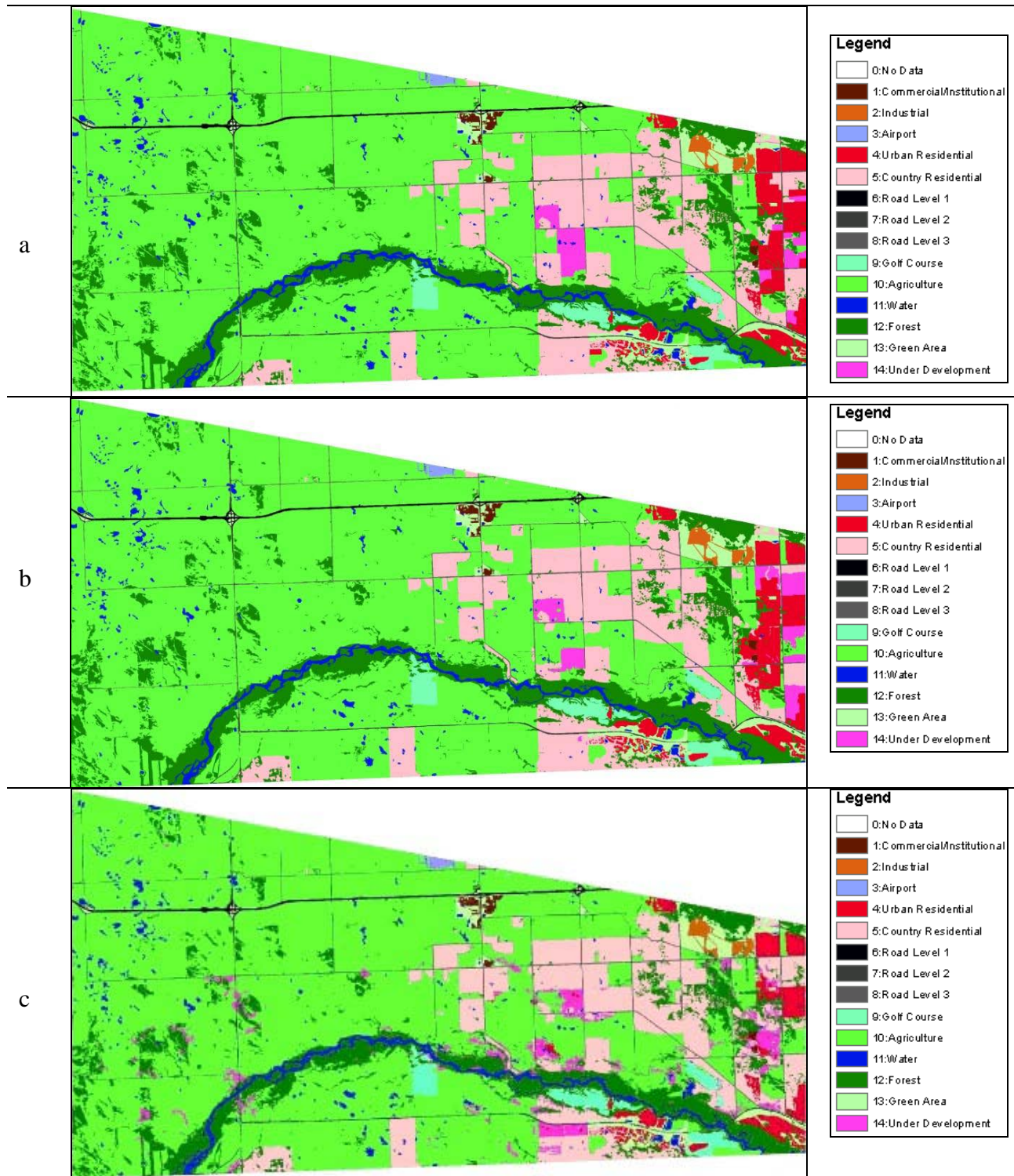


Figure 5-1 Comparison between historical land-use map (a) and the results generated with the patch-based (b) and the cell-based CA models (c) for the year 2011

Under Development is an important land-use class in the study area; the simulation pattern of Under Development will directly influence the simulated pattern of Country Residential and Urban Residential. Figure 5-2 shows the pattern of Under Development extracted from the historical land-use map of 2011, the simulated land-use map from the patch-based CA model, and the simulated land-use map from the cell-based CA model. We can observe that the pattern generated by the patch-based CA model is more compact and similar to the one present in the historical land-use maps.

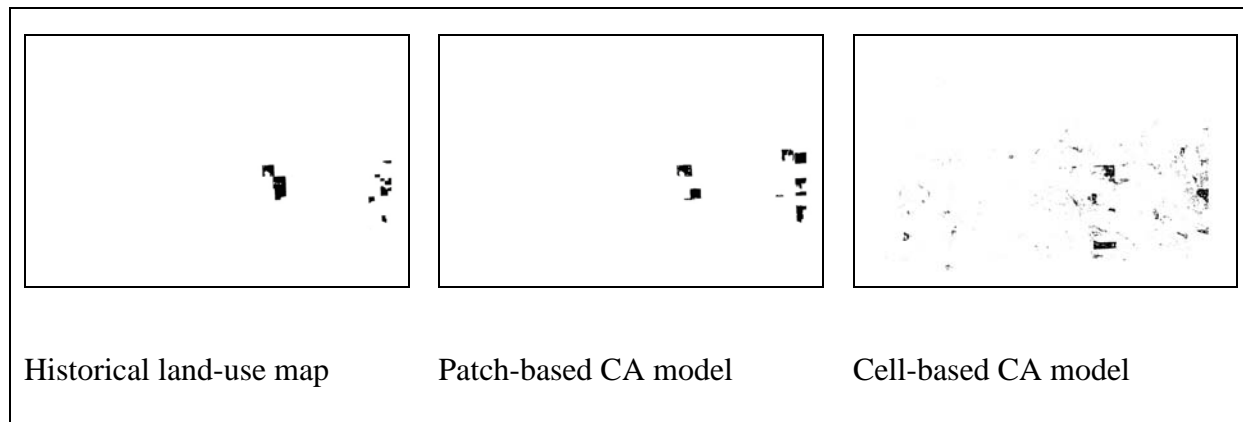


Figure 5-2 Comparison of the land-use class Under Development as observed in the historical land-use map and in the maps generated with the patch-based and the cell-based CA models for the year 2011

The $K_{simulation}$ values (Table 5-3) also reveal that the patch-based CA model generates patterns that are more similar to the reference land-use map than the ones produced by the conventional cell-based CA for the three years (2003, 2006, 2008) of the simulation. These values decrease over time for both models. However, for the patch-based model, the values are almost identical in 2006 (0.66) and 2008 (0.65), decreasing at 0.47 in 2011, while for the cell-based CA model, the values decrease from 0.60 in 2006, to 0.36 in 2008, becoming as low as 0.22 in 2011. This

means that the patch-based CA model shows a more stable performance when simulating land-use changes, while the performance of the cell-based CA model keeps decreasing. The reason is that the cell-based CA model introduces a great amount of sparsely distributed unrealistic patches/cells at each time step, which has a considerable influence on the simulation results for the final time step where errors have accumulated.

One important contribution of the patch-based CA is that it considerably improves the accuracy of location estimation. This can be observed from the high K_{Transloc} values (Table 5-3) generated by the patch-based CA model compared to the cell-based CA model, particularly for the years 2008 (0.84 vs. 0.46) and 2011 (0.71 vs. 0.32).

However, the estimated quantity of change as indicated by the $K_{\text{Transition}}$ values (Table 5-3) of the patch-based CA model is not ideal. Specifically for the years 2006 and 2011, these values are slightly higher for the outcomes generated by the cell-based CA than those of the patch-based CA (0.85 vs. 0.76 for the year 2006; 0.70 vs. 0.67 for the year 2011), while being the same for the year 2008 (0.78). This indicates that the patch-based model does not accurately estimate the quantity of land-use changes in the study area. Since the patches are typically much larger than individual cells, an error in the number of patches allocated to a specific land use by the patch-based CA model has a higher impact on the calculation of $K_{\text{Transition}}$ values than when using the cell-based CA. This will also be an issue when using the patch-based CA model for testing different scenarios in which the change between scenarios is small. For example, if scenario I indicates that the number of changed cells is 4520 at a time step and scenario II indicates that the number is 4000 cells, while the minimum patch size is 3000 cells and the mean patch size is

4000 cells, the patch-based CA model will generate one patch regardless of the difference between the number of changed cells in the two scenarios.

Table 5-3 $K_{Simulation}$ indices obtained for the patch-based and the cell-based CA models

Year	Index	Patch-based CA	Cell-based CA
2006	$K_{Simulation}$	0.66	0.60
	$K_{Transition}$	0.76	0.85
	$K_{Transloc}$	0.88	0.71
2008	$K_{Simulation}$	0.65	0.36
	$K_{Transition}$	0.78	0.78
	$K_{Transloc}$	0.84	0.46
2011	$K_{Simulation}$	0.47	0.22
	$K_{Transition}$	0.67	0.70
	$K_{Transloc}$	0.71	0.32

To analyze the land-use patterns that were generated by the models, landscape metrics were calculated on the three final land-use states (i.e., Under Development, Country Residential, and Urban Residential); the values are listed in Table 5-4. For the class Under Development, the number of patches generated by the patch-based CA model is considerably smaller than the one from the cell-based CA while being quite close to the one observed in the historical data for the three years of the simulation. The MPS and PSCOV generated by the patch-based CA model are also closer to the values calculated from the historical data. The patch-based CA produces considerably less variability in patch sizes (indicated by PSCOV) than the cell-based CA, with values that are in better accordance with the values calculated from the historical data. The patch-based CA model generates NOP values that are slightly smaller than those observed in the historical data and MPS values that are slightly larger than those in the historical data. These

results could be further improved by making the two parameters NOP and MPS adjustable in the model.

For the class Country Residential, both models overestimate the NOP (Table 5-4). The overestimations from the patch-based CA are almost constant for the three years (1.3 times for 2006, 1.4 times for 2008 and 1.2 times for 2011), while the overestimations from the cell-based CA increases over time (1.2 times for 2006, 1.7 times for 2008 and 3.8 times for 2011). While the MPS and PSCOV values calculated from the historical data remain almost identical through the three years of the simulation (about 15 for MPS and 311 for PSCOV), the MPS using the cell-based CA decreases over time (from 12.70 ha in 2006 to 0.17 ha in 2011) and the PSCOV increases (from 345.36 ha in 2006 to 607.86 ha in 2011). In comparison, the MPS and PSCOV values obtained with the patch-based model vary slightly over time and remain closer to the values found in the historical data.

For the class Urban Residential, the patch-based CA generates NOP and MPS values that are similar to the values obtained from the historical data, while the cell-based CA considerably overestimates the NOP (54 vs. 402) and underestimates MPS (13.67 vs. 0.05) and PSCOV (186.89 vs. 573.03), particularly for the year 2011 as indicated by these values.

Table 5-4 Landscape metrics (i.e., number of patches (NOP), mean patch size (MPS), and patch size coefficient of variation (PSCOV))

			Historical Data	Patch-based CA	Cell-based CA
2006	Under Development	NOP	19	13	416
		MPS	19.48	30.88	0.89
		PSCOV	127.10	136.27	662.29
	Country Residential	NOP	191	262	231
		MPS	15.36	11.07	12.70
		PSCOV	311.12	374.12	345.36
	Urban Residential	NOP	50	52	73
		MPS	9.94	9.20	6.81
		PSCOV	195.69	202.25	245.16
2008	Under Development	NOP	29	27	229
		MPS	14.22	16.89	1.86
		PSCOV	217.05	191.46	487.63
	Country Residential	NOP	187	264	327
		MPS	15.51	10.98	8.92
		PSCOV	310.75	375.74	417.07
	Urban Residential	NOP	52	53	63
		MPS	12.40	11.40	9.33
		PSCOV	180.96	186.90	202.49
2011	Under Development	NOP	35	27	354
		MPS	7.95	12.12	0.75
		PSCOV	289.94	168.51	964.59
	Country Residential	NOP	184	237	695
		MPS	16.49	12.20	3.17
		PSCOV	311.94	369.95	607.86
	Urban Residential	NOP	54	53	102
		MPS	13.67	14.06	4.05
		PSCOV	186.89	251.39	576.03

The introduction of the patch-based CA in this study successfully produced compact land-use patterns that are similar to those observed in the historical land-use maps. The performance of the patch-based CA model is higher when simulating the classes Urban Residential and Under

Development than Country Residential. This could be explained by the fact that fewer changes occurred in these two classes compared to Country Residential, but more importantly, the transitions to Under Development and Urban Residential happened in compact areas over a short period of time, a phenomenon that is well captured by the patch-based CA model. It can be observed from the NOP values found in the historical data that Urban Residential slightly increases and that Under Development expands over time at the expense of Country Residential. These trends are also reflected in the NOP values obtained with the patch-based CA, while the cell-based CA is not able to provide a similar interpretation of the land-use dynamics in the study area as it was expected.

To verify whether the patch-based CA model is successful in simulating long-term land-use changes, simulations were run for 30 years until 2041 using the initial land-use map of year 2011. The simulated result for the year 2041 is displayed in Figure 5-3. Again, the land-use patterns produced by the patch-based CA model are compact and consistent with those observed in the historical maps. Both Country Residential and Urban Residential expand from east to west. The expansion follows a typical space-filling process in which new urban land-use classes occur nearby existing ones. According to the simulation, by 2041, there will be an additional 59.5 km² of urban land, which will occupy 24.4% of the whole study area.

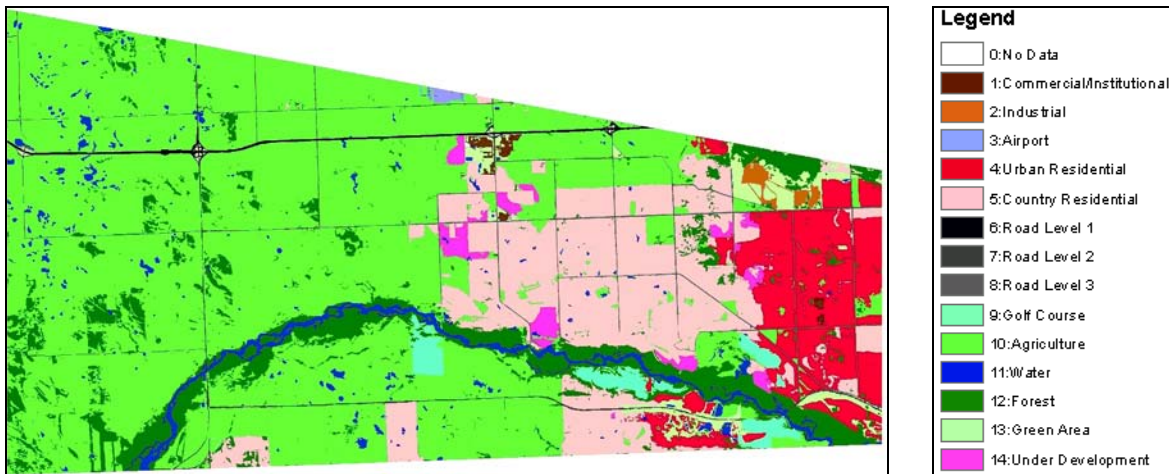


Figure 5-3 Simulated land-use map for the year 2041

5.2.2 Summary

When modeling urban land-use changes at fine spatial resolutions (e.g., 5m), traditional cell-based CA models tend to generate fragmented patterns because under this condition, the premise that a cell represents a real-world entity is rarely true. The fact is that one entity is usually composed of multiple cells that cannot be modeled separately. The entity-based CA models developed to date provide a more adequate representation of geographic space than the cell-based models. However, they neglect the internal heterogeneity of the entities that can be observed at fine resolutions. The patch-based CA model proposed in this study deals with this challenge by representing land-use entities using *patches* defined as a group of adjacent cells that might have different attributes but that represent a single entity. This patch-based CA model, along with a traditional cell-based CA were applied to simulate land-use changes in a dynamic area located at proximity of Calgary, a fast growing city in southern Alberta, Canada.

The performance of the patch-based CA model was examined using a visual comparison, $K_{simulation}$ indices, and landscape metrics (i.e., mean patch size, number of patches, and patch size

coefficient of variation). Compared to the cell-based CA model, the patch-based CA model generates more compact land-use patterns at a spatial resolution of 5 m, which is more consistent with the historical data. The $K_{\text{simulation}}$ indices indicate that the patch-based CA model generates land-use maps that are in greater agreement with the historical data than using the cell-based CA model. More importantly, the landscape metrics reveal that the patch-based CA is able to capture the most significant trends in the land-use dynamics as observed in the historical data, while the cell-based model is unable to provide a similar interpretation.

Traditional cell-based CA models are sensitive to the spatial resolution at which the landscape is represented and the transition rules defined (Ménard and Marceau, 2007; Pan et al., 2010). While the concept of patch tends to overcome this sensitivity by providing a landscape representation based on meaningful entities (rather than arbitrary cells), it is still expected that a change in spatial resolution will affect the simulation results. A further investigation of this issue could also be done.

The patch-based CA model serves as an appropriate solution to tackle the second challenge of space representation and simulation procedure for CA modeling at fine resolution. This CA model will then be used for evaluating various land development scenarios in Chapter 4.

5.3 Scenario Evaluation with the Patch-based CA Model

The simulated land-use maps under the three scenarios were acquired and analyzed for the simulation period of 2011-2041. The projected area for each land-use class for each time step was calculated and analyzed. Then the land-use maps of the final simulation year 2041 from the

three scenarios were compared. Finally, the consumption of non-developed land for each scenario was analyzed.

5.3.1 Projected area for each land-use class

The area of each dynamic land-use class at each year was used to analyze the change over time under three different scenarios.

For the class Country Residential, the growth trend is considerably different in the *business-as-usual scenario* than in the other two scenarios (Figure 5-4). While it keeps expanding in a steady rate in the *business-as-usual scenario* (an average of 0.87 km² every three years), it varies slightly in the two other scenarios (an average of 0.16 km² for the *protective growth scenario* and 0.15 km² for the *smart growth scenario*), even decreasing after the year 2035. The reason is that in these two scenarios, the development of Country Residential is limited. At the same time, large areas of Country Residential are converted to Under Development then to Urban Residential in the followings years. The net increase for Country Residential between 2011 and 2041 is 8.75 km², 1.64 km² and 1.51 km² for the three scenarios respectively, which indicates an increase of 30.3%, 5.7%, and 5.2% for the entire period.

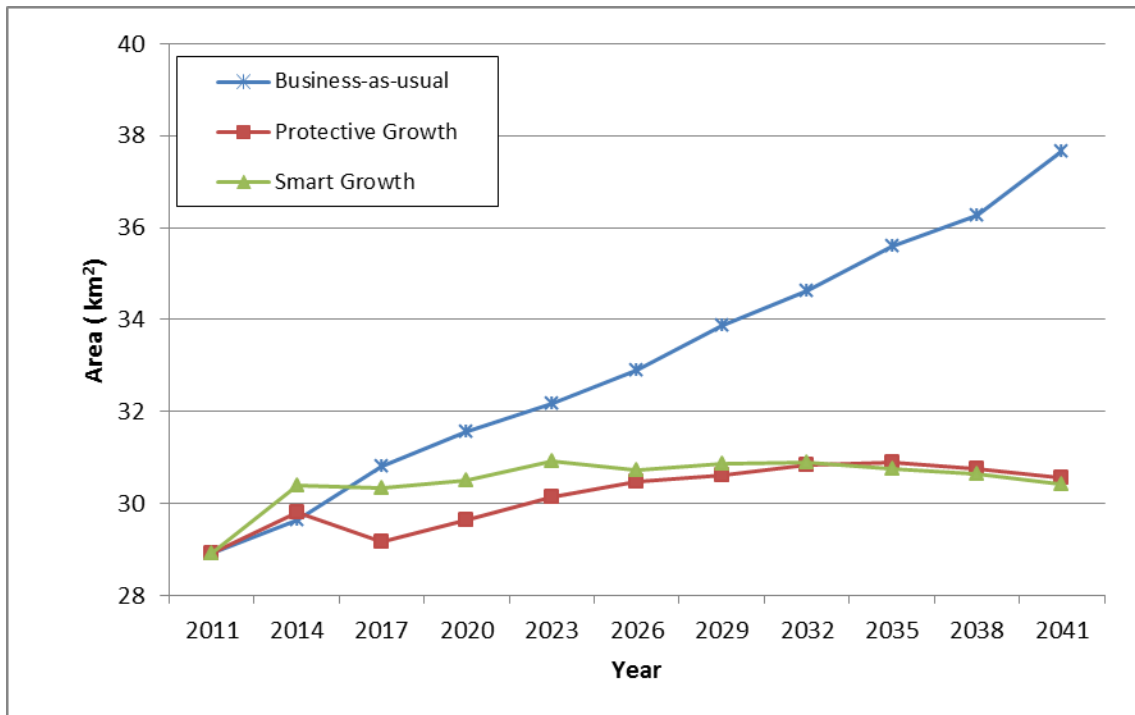


Figure 5-4 Projected area for country residential over the simulation period

For the class Urban Residential (Figure 5-5), a similar trend can be observed for the three scenarios. The area generated by the *business-as-usual scenario* (i.e., 11.24 km²) is close to the one generated by the *protective growth scenario* (i.e., 12.34 km²), while the area generated by the *smart growth scenario* is slightly smaller (i.e., 9.10 km²). Represented as percentage of increase, it is 151%, 166%, and 122% from 2011 for the three scenarios, respectively.

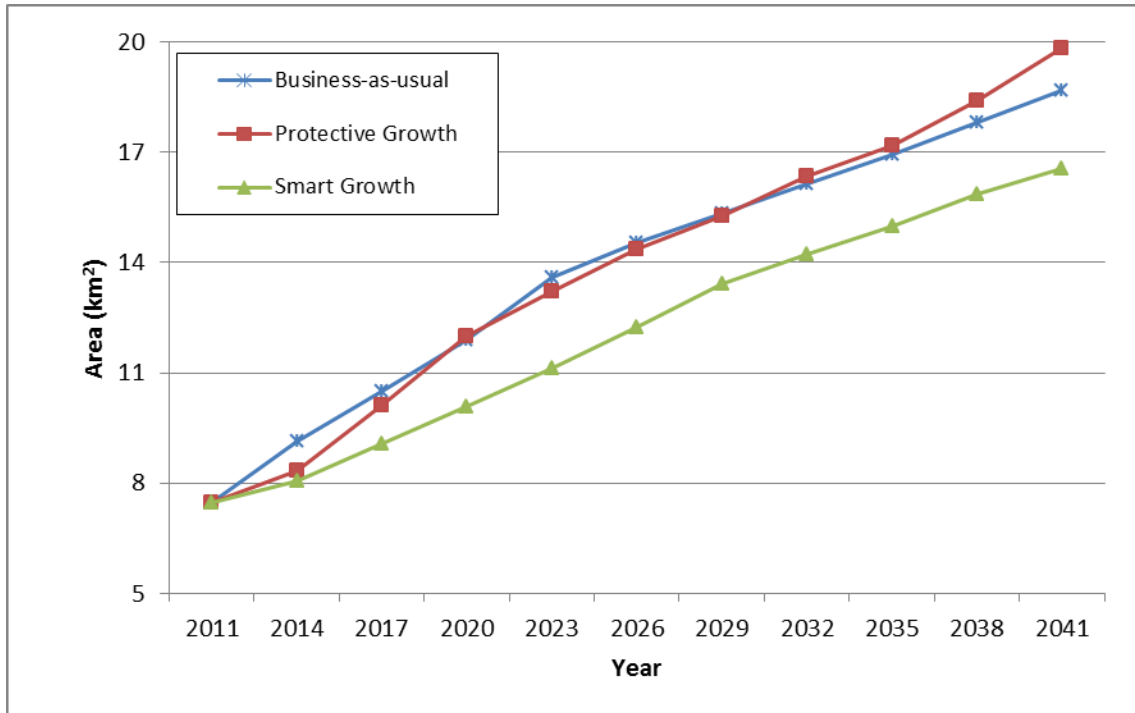


Figure 5-5 Projected area for urban residential over the simulation period

The agriculture area (Figure 5-6) declines in the three scenarios, but it is more pronounced in the *business-as-usual scenario* (an average of 1.63 km² every three years) and less important in the *smart growth scenario* (an average of 0.73 km² every three years). The difference between the decreased areas becomes larger over time for the three scenarios. From 2011 to 2041, the total agriculture area dropped is 16.3 km², 11.0 km², and 7.3 km² for the three scenarios respectively, which corresponds to 10.5%, 7.1%, and 4.7% of the total agriculture area in 2011.

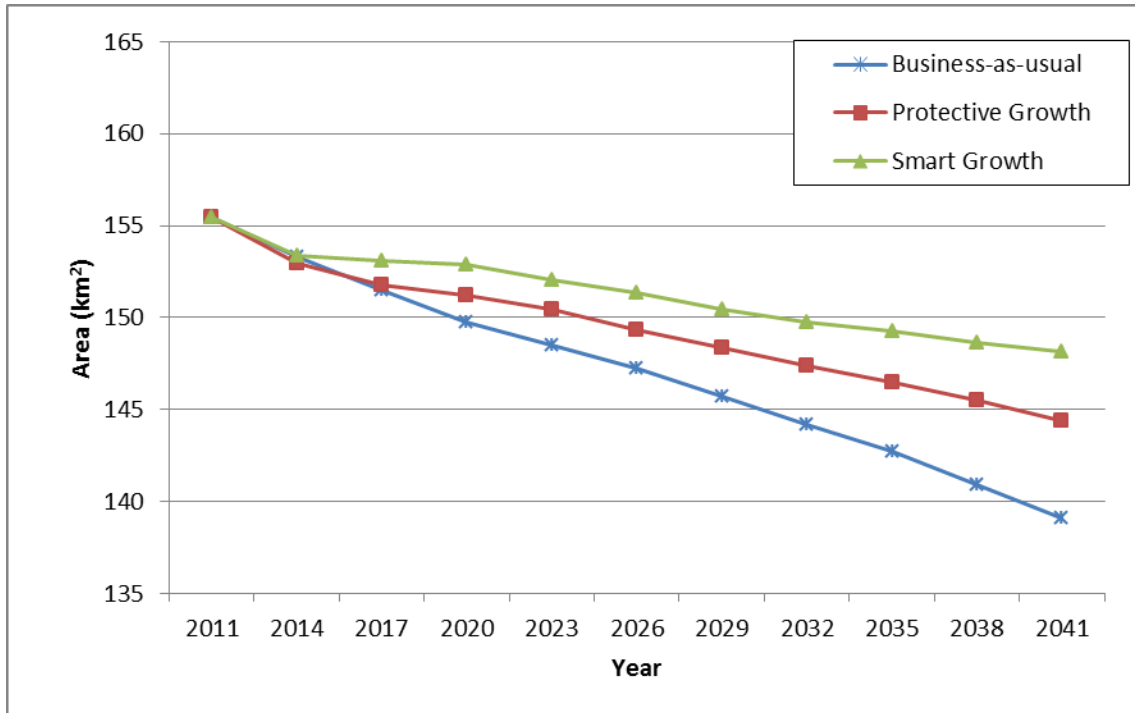


Figure 5-6 Projected area for agriculture over the simulation period

The decrease of forest area (Figure 5-7) is considerable in the *business-as-usual scenario* (i.e., 3.78 km² for the simulation period), while it is less pronounced in the two other scenarios (i.e., 1.71 km² in the *protective growth scenario* and 1.67 km² in the *smart growth scenario*). This suggests that the difference of the conversion from non-developed lands to developed lands between the protective growth and the *smart growth scenarios* is mainly generated by the amount of agriculture land consumption.

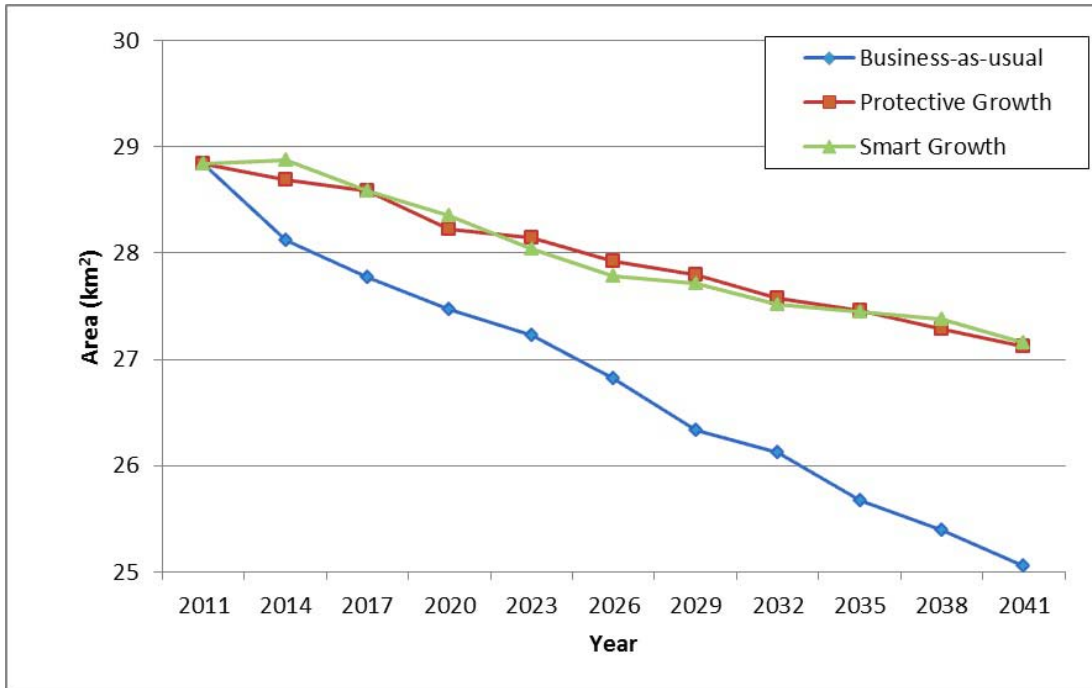


Figure 5-7 Projected area for forest over the simulation period

5.3.2 Visual comparison

The projected land-use maps for the year 2041 under the three scenarios are displayed in Figure 5-8. The first observation is that the land-use patterns produced by the CA model are compact and consistent with those observed in the historical maps.

The class Country Residential expands from east to west in the three scenarios. The expansion is more substantial in the *business-as-usual scenario* than in the two other scenarios, covering an area of 37.67 km² and resulting in an additional 7 km² of development (Table 5-5). In the *business-as-usual* and *smart growth scenarios*, the expansion of Country Residential tends to follow a space-filling pattern, where new developments occur around the existing Country Residential areas. With the incorporation of agriculture and water constraints in the *protective*

growth scenario, the development of Country Residential still appears around the existing Country Residential areas, but it avoids the good agriculture lands and lands that are too close to the river.

The expansion of the class Urban Residential occurs from east to west. In the *business-as-usual scenario*, almost all the non-developed lands within the city of Calgary boundary were converted to Urban Residential. In comparison, in the *protective growth scenario*, the good quality agriculture lands inside the Calgary limits was excluded and the growth of urban residential is therefore concentrated along the north transportation corridor (Highway 1). The *smart growth scenario* combines the trend of both the *business-as-usual* and the *protective growth scenario*, in which Urban Residential expands mostly from east to west, but is also concentrated along Highway 1. However the total area of Urban Residential generated in the *smart growth scenario* (i.e., 16.55 km²) is noticeably smaller than the one generated in the *business-as-usual scenario* (i.e., 18.69 km²) and the *protective growth scenario* (i.e., 19.83 km²) (Table 5-5).

In both the *business-as-usual* and *smart growth scenario*, the agriculture land close to the existing Country Residential and Urban Residential tends to be converted to urban land. While in the *protective growth scenario*, the good quality agriculture lands are kept untouched and the land development in proximity of the river is limited. The agriculture area in 2041 is 139.13 km², 144.42 km², 148.13 km² for the three scenarios respectively. These preserved areas, although relatively small in size, retain an important contribution to the ecosystem and agriculture in the study area.

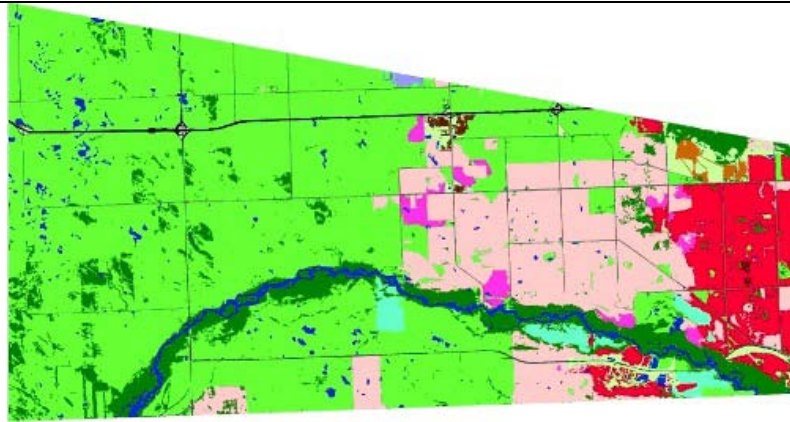
A large portion of forest located near Urban Residential in the *business-as-usual scenario* was converted into Urban Residential, while it was preserved in the other two scenarios. By 2041, the *protective growth* and *smart growth scenarios* are expected to preserve about 2 km² of forest comparing to the *business-as-usual scenario*. The preservation in the *protective growth scenario* basically occurs because these forested areas are close to the good quality agriculture lands, while the preservation in the *smart growth scenario* is mainly due to the lower requirement in Urban Residential area for the simulation period due to the constrained population growth. With the continuing urban growth in the following years, the preservation cannot be guaranteed if no further protection is involved.

The visual comparison suggests that the *business-as-usual scenario* is the least efficient in terms of land use, since it encourages the expansion of more urban land uses (i.e., Country Residential and Urban Residential), especially the low density Country Residential. The two other scenarios preserve more non-developed lands (i.e., Agriculture and Forest). In the *protective growth scenario*, the good quality agriculture lands were kept untouched and the growth close to the river was limited. These preserved areas, although relatively small in area, has an important contribution to the ecosystem and agriculture in the study area. This clearly demonstrates that incorporating development constrains and increasing land-use efficiency have a critical influence to both growth quantity and growth direction. The land-use maps generated by the patch-based CA model under the three scenarios illustrate that introducing constraints can better protect environmentally sensitive areas, while increasing land-use efficiency is an effective way to reduce the impact caused by a rapid population growth.

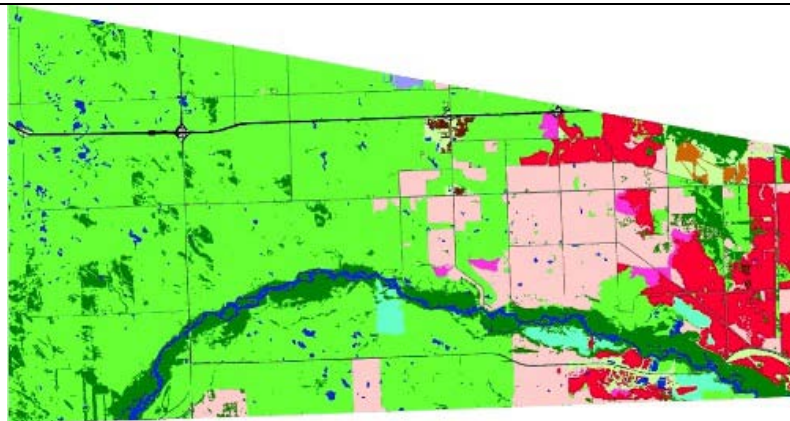
Projected land-use map of 2041

Legend

Business-as-usual Scenario



Protective Growth Scenario



Smart Growth Scenario

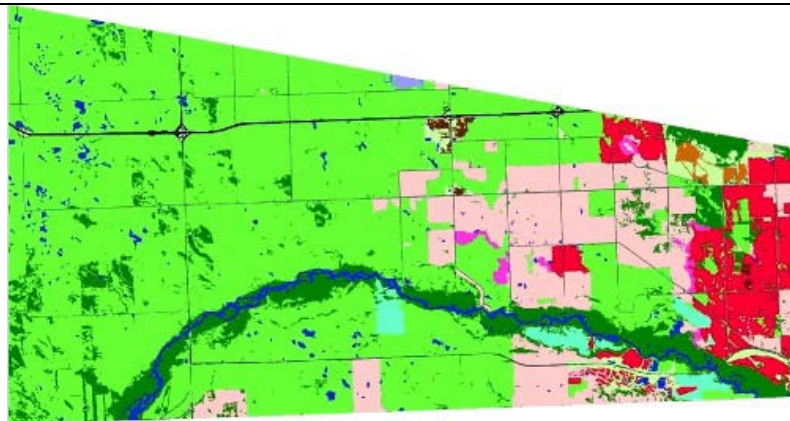


Figure 5-8 Projected land-use maps for the year 2041 under the three scenarios

Land composition statistics (Table 5-5) for the projected land-use maps at the year 2041 show that the *business-as-usual scenario* requires a total area of 59.5 km² of urban land, while the *protective growth scenario* only needs 52.14 km² of urban land by adopting higher land-use efficiency in order to accommodate the same size of population, which preserves 7.33 km² of non-developed land. By constraining the population, *smart growth scenario* only requires 48.38 km² of urban land in 2041. Therefore, urban settlement would occupy 24.4%, 21.4%, and 19.8% of the whole study area for the three scenarios respectively. This implies that by increasing the land-use efficiency by 30% in the *protective growth scenario*, the urban area will decrease by 12.3%. If population growth is further constrained to 2% each year, an additional 7.5% of land will be preserved.

Table 5-5 Land composition for the projected land-use maps under the three scenarios in 2041

Year	<i>Business-as-usual Scenario</i> (km ²)	<i>Protective Growth Scenario</i> (km ²)	<i>Smart Growth Scenario</i> (km ²)
Country Residential	37.67	30.56	30.43
Urban Residential	18.69	19.83	16.55
Agriculture	139.13	144.42	148.13
Forest	25.05	27.13	27.16

5.3.3 Land consumption

An important objective for land-use planning and scenario testing is to keep the land-use system sustainable over a long time period and limit the land consumption (i.e., limit the decrease in agriculture and forest). As noted in Table 5-6, the area consumed is larger for the year 2014 (i.e., 1.8% for the *business-as-usual scenario*, 1.7% for the *protective growth scenario*, and 1.3% for

the *smart growth scenario*) compared to the other years. For the remaining years, the consumption rate tends to be more stable, corresponding to 1.2% for the *business-as-usual scenario*, 0.7% for the *protective growth scenario*, and 0.5% for the *smart growth scenario*.

For the whole simulation period, as expected, the *business-as-usual scenario* is the least efficient in terms of land consumption with an area of 20.08 km² of non-developed land converted to developed land. Both the *protective growth* and *smart growth scenarios* consume less land (i.e., agriculture and forest) than the *business-as-usual scenario*. The *protective growth scenario* only consumes 12.71 km² of non-developed land in order to accommodate the same population size as the *business-as-usual scenario*, which preserves 7.37 km² of non-developed land. By constraining the population and adopting higher land-use efficiency, the *smart growth scenario* only converted 8.97 km² of non-developed land from 2011 to 2041. The percentage of land consumption is 12.8% for the *business-as-usual scenario*, 8.2% for the *protective growth scenario* and 5.6% for the *smart growth scenario*.

Table 5-6 Land consumption for each scenario over the simulation period

Year	<i>Business-as-usual Scenario</i>		<i>Protective Growth Scenario</i>		<i>Smart Growth Scenario</i>	
	Area (m ²)	Percentage (%)	Area (m ²)	Percentage (%)	Area (m ²)	Percentage (%)
2014	2.83	1.8	2.64	1.7	2.04	1.3
2017	2.18	1.4	1.25	0.8	0.53	0.3
2020	2.06	1.3	0.90	0.6	0.49	0.3
2023	1.44	0.9	0.88	0.6	1.15	0.7
2026	1.69	1.1	1.35	0.9	0.95	0.6
2029	1.99	1.3	1.07	0.7	0.98	0.6
2032	1.76	1.1	1.22	0.8	0.88	0.5
2035	1.90	1.2	0.98	0.6	0.57	0.4
2038	2.10	1.3	1.15	0.7	0.72	0.5
2041	2.13	1.4	1.26	0.8	0.70	0.4
Total	20.08	12.8	12.71	8.2	8.97	5.6

5.3.4 Summary

Using the information derived from various development plans related to the city of Calgary and Municipal District of Rocky View, the patch-based CA model was used to simulate land-use changes in the eastern Elbow River watershed under three predefined scenarios, namely, the *Business-as-usual scenario*, the *protective growth scenario*, and the *smart growth scenario*. The *Business-as-usual scenario* corresponds to a land development trend when all the conditions remain the same as that in the historical data. The *protective growth scenario* projects land development with the consideration of agriculture and water protection, as well as conserving more non-developed lands by increasing the land-use efficiency. The *smart growth scenario* utilizes the projected population in the study area to constrain the land development and increase the land-use efficiency at the same time. These scenarios were defined by introducing various

environmental constraints, incorporating population constraints, and modifying land-use efficiency, which characterized different growth strategies.

The three scenarios simulated with the patch-based CA model for the period 2011-2041 affect the main land-use classes differently. The total urban land areas (i.e., Country Residential and Urban Residential) generated by the three scenarios are different. There is an increase of 55% under the *Business-as-usual scenario*, 39% under the *protective growth scenario*, and 29% under the *smart growth scenario*. The *Business-as-usual scenario* generates a substantial increase of Country residential, mostly at the expense of agriculture and some forested areas, compared to a small increase in the two other scenarios. Urban residential increases for the three scenarios are similar, while Agriculture decreases are also about the same. Forest decreases sharply in the *Business-as-usual scenario*, and less so in the two other scenarios. Of the three scenarios, the *Business-as-usual scenario* tends to lead to the least efficient growth as expected, which consumes the highest percentage of non-developed land (12.8%), compared to the protective growth (8.2%), and smart growth (5.6%) scenarios. Smart growth maintains the largest agricultural area while maintaining the same forested area as the *protective growth scenario*.

From this study we also found that development of environmental sensitive areas can be limited by incorporating constraints. Additionally, the increase of land-use efficiency is proven to be an effective way of preserving more non-developed lands while accommodating a growing population. However, the patch-based CA model was only used to simulate the land-use changes in the study area for 30 years. Land-use dynamics in the study area over a longer period of time remains unexplored.

Chapter Six: Conclusion and Future Work

This research aimed at developing a cellular automata model to simulate land-use change at fine spatial resolutions. To achieve this goal, two key challenges – factor selection and geographic space representation – were tackled. To the best of my knowledge, this study is among the first attempts to evaluate the effectiveness of CA for simulating land-use changes at very fine resolutions.

The first challenge is associated with identifying the dominant factors that drive land-use dynamics when calibrating CA models, especially in the context where there are a large number of factors to be considered. The data mining technique Rough Set Theory (RST) (Pawlak, 1982) was proposed to facilitate factor reduction and selection in this study. RST was first tested in a subset of the Elbow River watershed using land-use maps at 30 m resolution with 5 land-use classes. The RST selected factors were then translated into transition rules and used for the calibration of the land-use CA model. Simulation outcomes obtained using the 18 original factors and a subset of factors identified by RST were compared to reference land-use maps. Some interesting results were revealed. First, factor selection using RST retains the original factors, which facilitates the interpretation of the land-use processes. Second, nine and eight factors were selected from the original set of 18 factors for the land-use classes vegetation and forest respectively, which not only greatly simplifies the model structure, but also suggested that different factors are responsible for different types of land-use changes. Finally, factors selected by RST tend to generate a higher agreement with reference land-use maps than the original group of 18 factors; they are better at capturing quantity information than location information.

The RST method was further used for the identification of dominant driving factors in a small area of the eastern Elbow River watershed, which involved a finer resolution (5 m) and a considerably larger number of land-use classes (14 land-use classes) and driving factors (42 factors).

The second challenge involved defining appropriate space representation in CA models required at fine spatial resolution. A novel patch-based CA model was proposed and implemented in which the real-world entities were represented as patches. In this patch-based CA model, the transition probabilities were calculated for each cell within each land-use patch to consider the internal heterogeneity of the entities being represented using a weight of evidence method and the factors identified based on the RST factor selection method. Then, land-use changes were simulated by employing a patch-based procedure based on transition probability maps. The potential changes were identified based on mean probability values for patches, rather than those of individual cells. A traditional cell-based CA model was also constructed for comparison. The results revealed that the introduction of the patch concept in the CA model generated compact land-use patterns that were similar to those observed in the historical land-use maps, while the cell-based CA model produced very fragmented results, as expected. Also, the patch-based CA model considerably improved the accuracy of land-use change location estimation and was able to capture the most significant trends in the land-use dynamics as observed in the historical data.

Using the information extracted from various municipal development plans, the patch-based CA model was applied to simulate three development scenarios in the eastern Elbow River watershed. The time span for the scenarios testing was from 2011 to 2041 with a time step of

three years. Three scenarios were evaluated, namely the *business-as-usual scenario*, the *protective growth scenario* and the *smart growth scenario*. The *business-as-usual scenario* simulated land uses assuming that the current land-use change rate remains unchanged and that the developmental conditions are similar to those observed from historical data. The *protective growth scenario* aimed at agriculture and water protection, as well as adopting higher land-use efficiency to preserve non-developed lands. In this scenario, a buffer around the Elbow River was drawn to limit land developments that are too close to the river and the consumption of good quality agricultural lands. Land-use efficiency, defined as the area of land consumed to accommodate a certain number of people, was also increased by 30%. The *smart growth scenarios* constrained the development using projected population growth and increased the land-use efficiency to encourage the formation of more compact land-use pattern.

Results reveal that the *business-as-usual scenario* tends to lead to the least efficient growth, which consumes the largest amount of non-developed land (i.e., 12.8% of the total area), while the *protective growth scenario* consumes 8.2% and the *smart growth scenario* consumes 5.6%. By incorporating agriculture and water constraints in the *protective growth scenario*, the good quality agriculture lands and lands close to the river were excluded from development. This suggests that introducing environmental constraints is an effective way to exclude the development of the environmental sensitive area, while the increase of land-use efficiency contributes in preserving more non-developed lands while accommodating the growing population. Moreover, the land-use patterns produced by the patch-based CA model are more compact and consistent with those observed in the historical maps.

6.1 Thesis contributions

The main contributions of this thesis are summarized as follows:

1) A novel RST method was introduced to facilitate the selection of the key driving factors for the calibration of the CA model. This technique derives the most important factors from an original set while minimizing the redundancy and retaining the original factors. The reduction of the driving factors not only greatly simplified the transition rules in the CA model, but also facilitated the interpretation of the land-use dynamics.

2) A novel patch-based CA model was designed in which the real-world entities (i.e., the detailed land-cover/land-use classes) are represented as patches and the transition probabilities are calculated for each cell within each patch. This method not only allowed the generation of compact land-use pattern, but also was able to account for the internal heterogeneity of the entities being represented. This model can aid to understand detailed land-use dynamics and act as a useful tool for testing various land development scenarios.

3) Using the patch-based CA model, three land development scenarios were designed by combining water and agriculture constraints, adjusting the land-use efficiency, as well as guiding the land development using projected population growth rate. The simulation results provided valuable insight into the land-use dynamics of the eastern Elbow River watershed.

6.2 Future work

The patch-based approach proposed in this study appears as a simple and valuable solution to adequately capture the internal heterogeneity of land-use classes and their transitions over time at fine spatial resolutions. However, future improvements are still needed.

Currently, the model still needs predefined input parameters (i.e., the ‘number of changed cells’ and ‘number of changed patches’), which makes forecasting difficult if this information is not available in the future. It will be helpful to define an automated simulation procedure based only on the transition probability maps. It is anticipated that the improved procedure will simplify the simulation, as well as provides the model with more flexibility and adaptability.

The expansion of the urban land-use classes (i.e., Under Development, Country Residential, and Urban Residential) generated by the patch-based CA model does not exactly mimic what can be observed in the real world. For example, in reality, when a parcel of land is selected for development, not only the area will be greater than a certain amount, but also, the land should follow a certain shape such as square or rectangular with a reasonable width. A parcel of land cannot be developed if the width is too narrow. Therefore, to make the simulation more realistic, future implementation should include a shape index model as an additional parameter to guide the land-use transitions.

In the patch-based CA model developed in this thesis, the change from three types of land uses to Under Development were combined together for considering the simultaneous occurrence of the these three types of land-use change. Equal ratio was used for the combination, which may

ignore the fact that different change ratios exist in reality. In the future, unequal development ratios of these three types of changes can be explored to consider this situation.

Future transportation networks and city boundary were not identified and included in the model. It can be observed from the historical data that the city boundary and transportation network influence the land-use transition to Urban Residential. Therefore, with the introduction of these datasets, the simulation to Urban Residential may be improved. Additionally, constraints such as environmentally sensitive areas, archaeological or historical sites were not included. If related data are available in the future, these components could be incorporated to improve the model calibration and simulation outcomes.

References

- Agterberg F.P. and G.F., Bonham-carter, 1990. Deriving weights of evidence from geoscience contour maps for the prediction of discrete events. *Proceedings of the 22nd APCOM Symposium*, Berlin, Germany, Sept. 17-21, Technical University of Berlin, 2: 381-395.
- Aguilera F. Valenzuela L.M. and A. Botequilha-Leital, 2011. Landscape metrics in the analysis of urban land-use patterns: A case study in a Spanish metropolitan area. *Landscape and Urban Planning*, 99: 226-238.
- Ahmad F., Hamdan A.B., and A.A. Bakar, 2007. Mining for interesting indicators using reduction approach. *Proceedings of the International Conference on Electrical Engineering and Informatics*, Institut Teknologi Bandung, Indonesia, June 17-19.
- Almeida C.M., Batty M., Monteiro A.M.V., Camara G., Soares-Filho B.S., Cerqueira G.C. and C.L. Pennachin, 2003. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, 27: 481-509.
- Almeida C.M., Gleriani J.M., Castejon E.F. and B.S. Soares-Filho, 2008. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22 (9): 943-963.
- Al-Ahmadi K., See L., Heppenstall A. and J. Hogg, 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecological Complexity*, 6: 80-101.
- Applied History Research Group (1997-2001). Calgary and Southern Alberta. Calgary, University of Calgary. 2009: Calgary's Growing Population chart.
- Antrop M., 2004. Landscape change and the urbanization process in Europe. *Landscape and Urban Planning*, 67: 9-26.

- Antrop M. and V. Van Eetvelde, 2000. Holistic aspects of suburban landscapes: visual image interpretation and landscape metrics, *Landscape and Urban Planning*, 50: 43–58.
- Arai K. and A. Basuki, 2010. A cellular automata based approach for prediction of hot mudflow disaster area. *Computational Science and Its Application-ICCSA 2010*: 87-98.
- Ashby, W.R. (1956) *An Introduction to Cybernetics*, Chapman & Hall, London.
- Batty M., 1997. Cellular automata and urban form: a primer. *Journal of the American Planning Association*, 63(2): 266-274.
- Batty, M., 2007. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press.
- Batty, M. and Y. Xie, 1994. From cells to cities. *Environment and Planning B*, 21(7): 531-548.
- Batty M. and Y. Xie, 1999. Self-organized criticality and urban development. *Discrete Dynamics in Nature and Society*, 3(2-3):109-124.
- Batty M., Xie Y. and Z. Sun, 1999. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems*, 23: 205-233.
- Benenson I., 2004. Agent-based modeling: from individual residential choice to urban residential dynamics. In *Spatially Integrated Social Science: Examples in Best Practice*, Edited by: Goodchild, D and Janelle, D. G. 67-94. Oxford: Oxford University Press.
- Benenson I. and P.M. Torrens, 2004. Geosimulation: Object-based modeling of urban phenomena. *Computers, Environment and Urban Systems*, 28, 1-8.
- Benenson I., Omer I. and E. Hatna, 2002. Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design*, 29: 491-512.

- Bennett, D.A., and McGinnis, D. 2008. Coupled and Complex: Human-Environment Interaction in the Greater Yellowstone Ecosystem, USA. *Geoforum*, 39:833-845.
- Berlekamp, E. R., J. H. Conway, and R. K. Guy, 1982. *Winning Ways for Your Mathematical Plays*. Volume 2. AK Peters.
- Brown D., Page S., Riolo R., Zellner M. and W. Rand, 2005. Path dependence and the validation of agent-based models of land use. *International Journal of Geographical Information Science*, 19(2): 153-174.
- Blaschke T., 2006. The role of the spatial dimension within the framework of sustainable landscape and natural capital. *Landscape and Urban Planning*, 75(3-4):198-226.
- Blaschke T. and J. Strobl, 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift für Geoinformationssysteme*, 14(6): 12-17.
- Bonham-carter G.F., Agterberg F.P. and D.F. Wright, 1989. Weights of evidence modeling: a new approach to mapping mineral potential: In *Statistical Applications in the Earth Sciences*, Eds. F.P. Agterberg and G.F. Bonham-carter, Geological Survey of Canada, Paper 89-9, p. 171-183.
- Bounfour A. and E.F. Lambin, 1999. How valuable is remotely sensed information? The case of tropical deforestation modeling. *Space Policy*, 15(3):149-158.
- Bryant C.R., 1982. *The Rural Real Estate Market: Geographic Patterns of Structure and Change in an Urban Fringe Environment*. Department of Geography, University of Waterloo, Canada, p. 153 Volume 18.

Canada Land Inventory, 2008.

<http://geogratiss.cgdi.gc.ca/geogratiss/en/option/select.do?id=CB6E057B-0D85-9B22-29DE-6351369A8B02>

Cao M. 2011. Implementation of a vector-based cellular automata model for simulating land-use changes. *19th International Conference on Geoinformatics*, June 24-26, Shanghai.

Caruso, G., Peeters, D., Cavailhes, J., and Rounsevell, M. 2007. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*, 37(5): 542-567.

Central Springbank Area Structure Plan, 2001. Department of Planning and Development, Municipal District of Rocky View NO.44. Access from online:
<http://www.rockyview.ca/LinkClick.aspx?fileticket=RKoWI2qipOo%3D&tabid=446&mid=858>.

Chen Y., Li X., Wang S. and X. Liu, 2012. Defining agents' behaviour based on urban economic theory to simulate complex urban residential dynamics. *International Journal of Geographic Information Science*, iFirst: 1-18.

Chomitz K.M. and D.A. Gray, 1996. Roads, land use and deforestation: a spatial model applied to Belize. *The World Bank Economic Review*, 10: 487-512.

City of Calgary, 2008. The City Calgary Population Growth 1998-2008. Calgary, City of Calgary.

City of Calgary, 2009. Population Picture. Calgary, City of Calgary.

Clarke, K. C. 1997. A self-modifying cellular automaton model of historical. *Environment and planning B*, (24): 247-261.

- Clarke K.C., and L.J. Gaydos, 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12(7): 699-714.
- Clarke K.C., 2008. Mapping and modelling land use change: An application of the SLEUTH model. In Pettit, C., Cartwright, W., Bishop, I., Lowell, K., Pullar, D. and Duncan, D. (Eds.), *Landscape Analysis and Visualisation: Spatial Models for Natural Resource Management and Planning* (pp. 353-366). Berlin Heidelberg: Springer.
- Conway T.M. and R.G. Lathrop, 2005. Alternative land use regulations and environmental impacts: assessing future land use in an urbanizing watershed. *Landscape and Urban Planning*, 70(1): 1-15.
- Cooper, Merrill. 2006. *Centre City Issues and Opportunities: Final Research Synthesis, 1999-2005*. Calgary: City of Calgary, Community and Neighbourhood Services, Social Research Unit.
- www.calgary.ca/docgallery/bu/cns/centre_city/centre_city_issues_and_opportunities.pdf
- Couclelis H., 1985. Cellular worlds: a framework for modeling micro-macro dynamics. *Environment and Planning A*, 17:585-596.
- Couclelis H., 1989. Macrostructure and microbehavior in a metropolitan area. *Environment and Planning B*, 16: 151-154.
- Couclelis H., 1997. From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B: Planning and Design*, 24: 165-174.

- Courage K., A. Masamu, A. Bongo and M. Munyaradzi, 2009. Rural sustainability under threat in Zimbabwe-simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. *Applied Geograph*, 29: 435-447.
- Crecine, J. P., 1964. A Time-Oriented Metropolitan Model for Spatial Location. *Technical Bulletin No. 6*, Department of City Planning, Pittsburgh.
- Deal B. and D. Schunk, 2004. Spatial dynamic modeling and urban land use transformation: a simulation approach to assessing the costs of urban sprawl. *Ecological Economics*, 51: 79-95.
- de Nijs T. and E. Pebesma, 2010. Estimating the influence of the neighborhood in the development of residential areas in the Netherlands. *Environment and Planning B: Planning and Design*, 37 (1): 21-41.
- Dietzel C. and K.C. Clarke, 2007. Toward Optimal Calibration of the SLEUTH Land Use Change Model. *Transactions in GIS*, 11(1): 29-45.
- Dickson B.G., Prather J.W., Xu Y., Hampton H.M., Aumack E.N. and T.D. Sisk, 2006. Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecology*, 21:747-761.
- Duro D.C., Franklin S.E. and M.G. Dubé, 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118: 259-272.
- East Springbank Area Structure Plan, 2010.
- http://www.calgary.ca/_layouts/cocis/DirectDownload.aspx?target=http%3A%2F%2Fw

ww.calgary.ca%2FPDA%2FLUPP%2FDocuments%2FPublications%2Ffeast-springbank-asp.pdf&noredirect=1&sf=1

Eloranta, J.W. 1997, Efficiency of Blasting vs. Crushing & Grinding. *Proceedings of the twenty-third conference of Explosives and Blasting Technique*, Las Vegas, Nevada, February 2-6, 1997. International Society of Explosives Engineers, Cleveland, Ohio.

Engelen G., White, R. and I. Uljee, 1997. Integrating constrained cellular automata models, GIS and decision support tools for urban planning and policy making. *Decision support systems in urban planning*, 125-155.

ERWP, 2010. Elbow River Watershed Partnership. <http://www.erwp.org/>.

ESRI, 2012. <http://www.esri.com>, Redlands, California, USA

ESRI, 2007. *ArcGIS help file*. USA, ESRI Press.

Exelis Visual Information Solutions, 2012. <http://www.exelisvis.com>.

Fang, S., Gertner, G.Z., Sun, Z., and A.A. Anderson, 2005. The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning*, 73(4):294-306.

Feng Y. and Y. Liu, 2012. An optimized cellular automata based on adaptive genetic algorithm for urban growth simulation. *Advances in Spatial Data Handling and GIS*. Lecture notes in Geoinformatics and Cartography, Part 2: 27-38.

Ferreira M.E., Ferreira L.G.J., Miziara F. and B.S. Soares-Filho, 2012. Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation. *Journal of Land Use Science*, iFirst: 1-19.

Foley J.A., DeFries R., Asner G.P., Barford C., Bonan G., Carpenter S.R., Chapin F.S., Coe M., Daily G.C., Gibbs H.K., Helkowski J.S., Holloway T., Howard E.A., Kucharik C.J.,

- Monfreda C., Patz J.A., Prentice I.C., Ramankutty N. and P.K. Snyder, 2005. Global consequences of land use. *Science*, 309(5734):570-574.
- Francis S.R. and J. Hamm, 2011. Looking forward: using scenario modeling to support regional land use planning in Northern Yukon, Canada. *Ecology and Society*, 16(4): 18.
- Fragkias M. and J. Geoghegan, 2010. Commercial and industrial land use change, job decentralization and growth controls: a spatially explicit analysis. *Journal of Land Use Science*, 5(1): 45-66.
- Fu B. and L. Chen, 2000. Agricultural landscape spatial pattern analysis in the semi-arid hill area of the Loess Plateau, China. *Journal of Arid Environments*, 44: 291-303.
- García A.M., Santé I., Boullón M. and R. Crecente, 2012. A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain. *Computers, Environment and Urban Systems*, 36(4):291-301.
- Gardner M., 1970. The fantastic combinations of John Conway's new solitaire game 'life'. *Scientific American*, 223(4): 120-123.
- Gardner M., 1971. On cellular automata, self-reproduction, the Garden of Eden and the game of 'life'. *Scientific American*, 224(2): 112-117.
- Gerling D., 1990. Natural enemies of whiteflies predators and parasitoids. D. Gerling (Ed.), *Whiteflies: their Bionomics, Pest Status and Management, Intercept*. Andover, Hants, UK: 147-185.
- GLP, 2005. Science Plan and Implementation Strategy. *IGBP Report No. 53/IHDP Report No. 19*, Stockholm, 64.

- Guan D., Li H., Inohae T., Su W., Nagaie T. and K. Hokao, 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecological Modelling*, 222: 3761-3772.
- Gutman, G. et al., 2004. *Land change science: Observing, monitoring and understanding trajectories of change on the earth's surface with remote sensing and digital image processing*. Kluwer Academic Publishers, 459 pp.
- Haerberlin Y., P. Turberg, A. Retiere, O. Senegas and A. Parriaux, 2004. Validation of SPOT-5 satellite imagery for geological hazard identification and risk assessment for landslides, mud and debris flows in Matagalpa, Nicaragua.
<http://cartesia.org/gedoc/isprs2004/comm/papers/51.pdf>,2004.
- Hagen-Zanker A. and G. Lajoie, 2008. Neutral models of landscape change as benchmarks in the assessment of model performance. *Landscape and Urban planning*, 86(3): 284-296.
- Hagoort, M. Geertman, S. and H. Ottens, 2008. Spatial externalities, neighbourhood rules and CA land-use modelling. *The Annals of Regional Science*, 42:39–56.
- Hamman Y., A. Moore and P. Whigham, 2007. The dynamic geometry of geographical vector agents. *Computers, Environment and Urban Systems*, 31: 502-519.
- Han J.C., Hu X.H. and C. Nick, 2001. Supervised learning: A generalized rough set approach. In Ziarko W. and Y. Yao (Eds.), *Rough Sets and Current Trends in Computing* (pp. 322-329). Heidelberg: Springer.
- Harbor J.M., 1994. A practical method for estimating the impact of land-use change on surface runoff, groundwater recharge and wetland hydrology. *Journal of American Planning Association*, 60(1): 95-108.

- Hasbani J.G., Wijesekara N. and D.J. Marceau, 2011. An interactive method to dynamically create transition rules in a land-use Cellular Automata model, *Cellular Automata - Simplicity Behind Complexity*, Alejandro Salcido (Ed.), ISBN: 978-953-307-230-2, InTech, Available from: <http://www.intechopen.com/articles/show/title/an-interactive-method-to-dynamically-create-transition-rules-in-a-land-use-cellular-automata>.
- Hassen Y. and E.Tazaki, 2003. Adaptive behavior in cellular automata using rough set theory. *Applied Artificial Intelligence*, 17(2): 155-175.
- He C., Okada N., Zhang Q., Shi P. and J. Zhang, 2006. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography*, 26(3-4): 323-345.
- Hu S. and D. Li, 2004. Vector cellular automata based geographical entity. *Proceeding of the 12th International Conference on Geoinformatics*, June 7-9, Sweden.
- Hoggart, K. (Ed.) 2005. *The City's Hinterland. Dynamism and Divergence in Europe's Peri-urban Territories*, pp. 1-18 (Aldershot: Ashgate).
- Houet T., Loveland TR., Hubert-Moy L., Gaucherel C. Napton D., Barnes CA. and K. Sayler, 2010. Exploring subtle land use and land cover changes: a framework for future landscape studies. *Landscape Ecology*, 25(2):249-266.
- IDP, 2011. *Rocky View County/City of Calgary Intermunicipal Development Plan*. Access from online: http://www.municipalink.ca/reports/docs/rocky_view_idp_june_2011.pdf
- Iltanen S., 2012. Cellular automata in urban spatial modelling. *Agent-based Models of Geographical Systems*, 69-84.
- Imagine Calgary Plan for Long Range Urban Sustainability, 2006.
http://www.imaginecalgary.ca/imagineCALGARY_long_range_plan.pdf

- Irwin, E.G., 2010. New directions for urban economic models of land use change: incorporating spatial dynamics and heterogeneity. *Journal of Regional Science*, 50: 65-91.
- Irwin E.G. and N.E. Bockstael, 2002. Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*, 2:31-54.
- Jantz, C.A., Goetz, S.J., and M.K. Shelley, 2004. 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*, 31(2): 251-272.
- Jantz C.A., Goetz S.J., Donato D. and P. Claggett, 2010. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*, 34(1):1-16.
- 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.
- Jensen, R., & Shen, Q. (2007). Rough set based feature selection: A review. *Rough Computing: Theories, Technologies and Applications*. IGI global.
- Kok K. and A. Veldkamp, 2001. Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture, Ecosystems and Environment*, 85: 205-221.
- Kok K., Farrow A., Veldkamp A. and P.H. Verburg, 2001. A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems and Environment*, 85(1-3):223-238.
- Kressler F.P., Kim Y.S. and K.T., Steinnocher, 2003. Object-oriented land cover classification of panchromatic KOMPSAT-1 and SPOT-5 data, *Proceedings of IGARSS 2003 IEEE*, Toulouse.

- Kueppers, L., Baer, P., Harte, J., Haya, B., Koteen, L., and M. Smith, 2004. A decision matrix approach to evaluating the impacts of land-use activities undertaken to mitigate climate change. *Climatic Change*, 63(3):247-257.
- Lambin E.F., Rounsevell M.D.A. and H.J. Geist, 2000. Are agricultural land-use models able to predict changes in land-use intensity? *Agriculture, Ecosystems and Environment*, 82(1-3):321-331.
- Landis J., 1994. The California Urban Futures Model: A New Generation of Metropolitan Simulation Models. *Environment and Planning B: Planning and Design*, 21:399-420.
- Lau K.H. and B.H. Kam, 2005. A cellular automata model for urban land-use simulation. *Environment and Planning B*, 32: 247-263.
- Li X., 2011. Emergence of bottom-up models as a tool for landscape simulation and planning. *Landscape and Urban Planning*, 100: 393-395.
- Li X. and A.G.-O. Yeh, 2000. Modelling sustainable urban development by the integration of constrained cellular automata. *International Journal of Geographical Information Science*, 14(2):131-152.
- Li X. and A.G.-O Yeh, 2001. Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environment and Planning A*, 33(8):1445-1462.
- Li X. and A.G-O Yeh, 2002. Integration of principal components analysis and cellular automata for spatial decision making and urban simulation. *Science in China*, 45(6), 521-529.
- 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 H. and H. Motoda (Eds.), 1998. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, Boston.

- Liu X., Li X., Shi X., Zhang X. and Y. Chen, 2010. Simulating land-use dynamics under planning policies by integrating artificial immune systems with cellular automata. *International Journal of Geographical Information Science*, 24(5): 783-802.
- Liu Y. and R. Phinn, 2003. Modeling urban development with cellular automata incorporating fuzzy-set approaches. *Computers, Environment and Urban Systems*, 27:637-658.
- Long Y., Shen Z. and Q. Mao, 2012. Retrieving spatial policy parameters from an alternative plan using constrained cellular automata and regionalized sensitivity analysis. *Environment and Planning B*, 39(3): 586.
- Lösch A., 1940. *The Economics of Location*, Jena : Fisher. English translation, New Haven: Yale University Press, 1954.
- Lowry I.S., 1964. *A Model of Metropolis*. Santa Monica: Rand Corporation.
- Lyford M.E., Jackson, S.T., Betancourt, J.L. and S.T. Gray, 2003. Influence of landscape structure and climate variability on a late Holocene plant migration. *Ecological Monographs*, 77: 567-583.
- McGarigal K, Marks B, Ene E and C. Holmes, 1995. Fragstats. Department of Natural Resources Conservation, University of Massachusetts, Amherst, MA, <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.
- Meeus S. and H. Gulinck, 2008. Semi-Urban Areas in Landscape Research: A Review. *Living Reviews in Landscape Research*, 2(3):1-45.
- 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.

- Mitsova D., Shuster W. and X. Wang, 2011. A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landscape and Urban Planning*, 99(2):141-153.
- Moore A., 2011. Geographical vector agent-based simulation for agricultural land-use modelling. In *Advanced Geosimulation Models*, D.J. Marceau and I. Benenson, eds, Bentham Science Publisher, pp. 30-48.
- Moreno N., Menard A. and D.J. Marceau, 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design*, 35: 647-665.
- Moreno N., F. Wang and D.J. Marceau, 2009. Implementation of a dynamic neighbourhood in a land-use vector-based geographic cellular automata model. *Computers, Environment and Urban Systems*, 33: 44-54.
- Moreno N., F. Wang and D.J. Marceau, 2010. A geographic object-based approach in cellular automata modeling. *Photogrammetric Engineering and Remote Sensing*, 76(2): 183-191.
- Myint S.W. and L. Wang, 2006. Multicriteria decision approach for land use land cover change using Markov chain analysis and a cellular automata approach. *Canadian Journal of Remote Sensing*, 32: 390-404.
- Municipal District of Rocky View No.44, 1998. Municipal Development Plan. Accessed from online:
<http://www.rockyview.ca/LinkClick.aspx?fileticket=YbEP3GoBST0%3D&tabid=195>.
- Nguyen H.S., 1998. Discretization problem for Rough Sets Methods. In: L. Polkowski and A. Skowron (Eds.), *Rough Sets and Current Trend in Computing* (pp. 545-552) (*Proceedings of the First International Conference, RSCTC'98*), June, Warsaw, Poland.

- Norman L.M., Feller M. and M.L. Villarreal, 2012. Developing spatially explicit footprints of plausible land-use scenarios in the Santa Cruz Watershed, Arizona and Sonora. *Landscape and Urban Planning*, 107(3): 225-235.
- Øhrn A., Komorowski J., Skowron A., and P. Synak, 1998. The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets: The ROSETTA System. In L. Polkowski and A. Skowron (Eds.), *Rough Sets in Knowledge Discovery 1: Methodology and Applications, Studies in Fuzziness and Soft Computing* (pp. 376-399), Physica-Verlag.
- Øhrn A., 1999. *Discernibility and rough sets in medicine: Tools and applications*. Ph.D. thesis, Department of Computer and Information Science, Norwegian University of Science and Technology.
- Onsted J.A. and K.C. Clarke, 2011. Forecasting enrolment in differential assessment programs using cellular automata. *Environment and Planning B: Planning and Design*, 38: 829-849.
- Paelinck J., 1970. Dynamic urban growth models. *Papers in Regional Science*, 24:25-38.
- Pan Y., Roth A., Yu Z. and D. Reiner, 2010. The impact of variation in scale on the behavior of a cellular automata used for land-use change modeling. *Computers, Environment and Urban System*, 34(5): 400-408.
- Pang, M.Y.C. and W.Z. Shi (2002), Development of a process-based model for dynamic interaction in spatio-temporal GIS. *Geoinformatica*, 6:323-344.
- Parks Foundation Calgary, 2008. At home on the Elbow River. Accessed from online: <http://www.parksfdn.com/At%20Home%20on%20the%20Elbow%20River.pdf>.
- Parker, D.C., Manson S.M., Janssen M. A., Hoffmann M.J. and P. Deadman, 2003. Multi-agent systems for the simulation of land-use and land- cover change: A review. *Annals of the*

- Association of American Geographers*, 93(2): 314-337.
- Pawlak, Z. 1982. Rough sets. *International Journal of Parallel Programming*, 11(5): 341-356.
- Pawlak Z., 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Pawlak Z., 1998. Rough set theory and its applications to data analysis. *Cybernetics and Systems: An International Journal*, 29: 661-688.
- Pérez-Vega, Mas J.-F. and A. Ligmann-Zielinska, 2012. Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. *Environmental Modelling and Software*, 29(1): 11-23.
- Peng J., Wang Y., Zhang Y., Wu J., Li W. and Y. Li, 2010. Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. *Ecological Indicators*, 10: 217-223.
- Perraton, J., and Baxter, R. 1974. *Models, Evaluations & Information Systems for Planners*. MTP Construction, Lancaster, England.
- Phipps M., 1989. Dynamical behavior of cellular automata under constraints of neighbourhood coherence. *Geographical Analysis*, 21:197-215.
- Phipps M., 1992. From local to global: the lesson of cellular automata. D.L. DeAngelis, L.J. Gross (Eds.), *Individual Based Models and Approaches in Ecology: Populations, Communities and Ecosystems*, Chapman & Hall, New York, pp. 165-187.
- Pinto N.N. and A.P. Antunes, 2010. A cellular automata model based on irregular cells: application to small urban areas. *Environment and Planning B*, 37: 1095-1114.
- Plan it Calgary, 2007. *Plan it Calgary Workbook*. Sustainable City Program, City of Calgary.
- Access from online:
http://www.calgary.ca/docgallery/BU/planning/pdf/plan_it/plan_it_workbook_final.pdf

- Pohl C. and J.L. Van Genderen, 1998. Review article mu Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5):823-854.
- Polhill J.G., Parker D. and N.M. Gotts, 2008. Effects of land markets on competition between innovators and imitators in land use: results from FEARLUS-ELMM. *Social simulation technologies: Advances and new discoveries*, 81-97.
- Portugali J. and I. Benenson, 1995. Artificial planning experience by means of a heuristic cell-space model: simulating international migration in the urban process. *Environment and Planning A*, 27: 1647-1647.
- Portugali J., Benenson I. and I. Omer, 1994. Sociospatial Residential Dynamics: Stability and Instability within a Self- Organizing City. *Geographical Analysis*, 26(4):321-340.
- Questier F., Arnaut-Rollier I., Walczak B. and D.L. Massart, 2002. Application of rough set theory to feature selection for unsupervised clustering. *Chemometrics and Intelligent Laboratory Systems*, 63:155-167.
- Raines G.L. and M.J. Mihalasky, 2002. A reconnaissance method for regional-scale mineral-resource assessment, based exclusively on geologic-map data. *Natural Resources Research*, 11(4): 241-248.
- RIKS BV (2012). <http://www.riks.nl/mck>.
- Robinson J.G., 2011. Ethical pluralism, pragmatism, and sustainability in conservation practice. *Biological Conservation*, 144(3):958-965.
- Rosetta development team, 2009. <http://www.lcb.uu.se/tools/rosetta>.

- Samat N., 2006. Characterizing the scale sensitivity of the cellular automata simulated urban growth: A case study of the Seberang Perai Region, Penang State, Malaysia. *Computers, Environment and Urban Systems*, 30(6):905-920.
- Santé I., Garcia A.M., Miranda D. and R. Crecente, 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2):108-122.
- Schmit C., Rounsevell M.D.A. and I.L. Jeunesse, 2006. The limitations of spatial land use data in environmental analysis. *Environmental Science and Policy*, 9(2):174-188.
- Serneels S. and E.F. Lambin, 2002. Impact of land-use changes on the wildebeest (*Connochaetes taurinus*) in the northern part of the Serengeti-Mara ecosystem. *Journal of Biogeography*, 28 (3):391-408.
- Shan J., Alkheder S. and J. Wang, 2008. Genetic algorithms for the calibration of cellular automata urban growth modeling. *Photogrammetric Engineering and Remote Sensing*, 74(10): 1267-1277.
- Shen Q., Chen Q., Tang B., Yeung S., Hu Y. and G. Cheung, 2009. A system dynamics model for the sustainable land use planning and development. *Habitat International*, 33(1):15-25.
- Silva E. and N. Wu, 2012. Surveying models in urban land studies. *Journal of Planning Literature*, 27(2):139-152.
- Silva E. A. and K. C. Clarke, 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26(6): 525-552.

- Soares-Filho B.S., Coutinho Cerqueira G. and C. Lopes Pennachin, 2002. DINAMICA-a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological modelling*, 154(3), 217-235.
- Solecki W.D. and C. Oliveri, 2004. Downscaling climate change scenarios in an urban land use change model. *Journal of Environment Management*, 72:105-115.
- Sosiak A and J. Dixon, 2004. Impacts on water quality on the Upper Elbow River. Alberta Environment and City of Calgary, Calgary.
- Spiegelhalter DJ., 1986. Probabilistic prediction in patient management and clinical trials. *Statistics in Medicine*, 5: 421-433.
- Stanilov K., 2003. Accessibility and land use: the case of suburban Seattle, 1960-1990. *Regional Studies*, 37(8), 783-794.
- Stanilov K. and M. Batty, 2011. Exploring the historical determinants of urban growth patterns through Cellular Automata. *Transactions in GIS*, 15(3): 253-271.
- Statistics Canada, 2009. Annual population estimates and demographic factors of growth by census metropolitan area, Canada, from July to June — Population Estimates and Factors of Growth. Statistics Canada. July 1, 2009. <http://www.statcan.gc.ca/pub/91-214-x/2008000/t021-eng.htm>.
- Statistics Canada, 2012.
<http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=302&PR=48&S=51&O=A&RPP=25>.
- Stevens D., Dragicevic S. and K. Rothley, 2007. iCity: a GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling and Software*, 22: 761-773.

- Stevens D. and S. Dragicevic, 2007. A GIS-based irregular cellular automata model of land-use change. *Environment Planning B*, 34: 708-724.
- Suarez-Rubio M., Lookingbill T.R. and L.A. Wainger, 2012. Modeling exurban development near Washington, DC, USA: comparison of a pattern-based model and a spatially-explicit econometric model. *Landscape Ecology*, 27(7): 1045-1061.
- 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: 37-52.
- Swiniarski R. and A. Skowron, 2003. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 24(6):833-849.
- 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.
- Tattoni C., Ciolli M. and F. Ferretti, 2011. The fate of priority areas for conservation in protected areas: a fine-scale Markov chain approach. *Environmental Management*, 47:263-278.
- Thapa R.B. and Y. Murayama, 2011. Urban growth modeling of Kathmandu metropolitan region, Nepal. *Computers, Environment and Urban systems*, 35(1): 25-34.
- Thapa R.B. and Y. Murayama, 2012. Scenario based urban growth allocation in Kathmandu Valley, Nepal. *Landscape and Urban Planning*, 105(1-2): 140-148.
- Tobler W., 1979. Cellular geography. In S. Gale, and G. Olsson (Eds.), *Philosophy in Geography* (pp. 379-386). Riedel, Dordrecht.
- Torrens P.M., 2006. Remote sensing as dataware for human settlement simulation". In *Remote Sensing of Human Settlements*, M. Ridd and Hipple, J. D.(Eds.). Bethesda: American Society of Photogrammetry and Remote Sensing, pp. 693-699.

- Torrens P.M. and D. O'Sullivan, 2001. Cellular automata and urban simulation: where do we go from here? *Environment and Planning B*, 28: 163-168.
- Torrens P.M. and I. Benenson, 2005. Geographic Automata Systems. *International Journal of Geographical Information Science*, 19(4): 385-412.
- Trimble, 2012. <http://www.ecognition.com>
- Van Vliet J., Bregt A. K. and A. Hagen-Zanker, 2011. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, 222(8): 1367-1375.
- Verburg, P. H., 2006. Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, 21(8): 1171-1183.
- Verburg P.H., de Nijs T.C.M., van Eck J.R., Visser H. and K. de Jong, 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28: 667-690.
- Verburg P.H., Koomen E., Hilferink M., Pérez-Soba M. and J.P. Lesschen, 2012. An assessment of impact of climate adaption measures to reduce flood risk on ecosystem services. *Landscape Ecology*, 27(4): 473-486.
- Verburg P.H. and K. Overmars, 2009. Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecology*, 24:1167-1181.
- Verburg PH, Veldkamp A, de Koning GHJ, Kok K and J. Bouma, 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modeling*, 116:45-61.

- Verburg P.H. and A. Veldkamp, 2001. The roles of spatially explicit models in land use change research sequences – a case study for cropping patterns in China. *Agriculture, Ecosystem and Environment*, 85:177-190.
- Verburg, P. H., P. P. Schot, M. Dijst and A. Veldkamp, 2004. Land use change modeling: current practice and research priorities. *GeoJournal*, 61: 309-324.
- Veldkamp A. and E.F.Lambin, 2001. Predicting land-use change. *Agriculture, Ecosystems and Environment*, 85(1-3):1-6.
- Veldkamp A. and L.O. Fresco, 1996. CLUE: a conceptual model to study the conversion of land use and its effects. *Ecological Modelling*, 85(2-3): 253-270.
- Veldkamp A., 2009. Investigating land dynamics: future research perspectives. *Journal of Land Use Science* 4(1-2): 5-14.
- Vinterbo S. and A. Øhrn, 2000. Minimal approximate hitting sets and rule templates. *International Journal of Approximate Reasoning*, 25(2): 123-143.
- Von Neumann J., 1963. General and logical theory of Automata. In, A.H. Taub (Ed.), *Collected Work 5* (pp. 288-328). Oxford: Pergamon Press.
- Wainger LA, Rayburn J and E.L. Price, 2007. Review of land use change models: applicability to projections of future energy demand in the Southeast United States. *University of Maryland Center for Environmental Science Technical Report Ref. No. 07-187*.
- Wang F., 2005. On acquiring classification knowledge from noisy data based on rough set. *Expert Systems with Applications*, 29: 49-64.
- Wang F., Hasbani J-G., Wang X., and D. J. Marceau, 2011. Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set theory. *Computers, Environment and Urban Systems*, 35(2):116-125.

- Wang S.Q., Zheng X.Q. and X.B. Zang, 2012. Accuracy assessments of land use change simulation based on Markov-cellular automata model. *Procedia Environmental Sciences*, 13: 1238-1245.
- Weber A., 1909. *The location of industries* (English edition, 1929). Chicago: University of Chicago Press.
- Wegener M., 1986. Transport network equilibrium and regional deconcentration. *Environment and Planning A*, 18: 437-456.
- Weins J.A., 1976. Population responses to patchy environments. *Annual Review of Ecology and Systematics*, 7: 81-120.
- Wen W., Wang W.K. and T.H. Wang, 2005. A hybrid knowledge-based decision support system for enterprise mergers and acquisitions. *Expert Systems with Applications*, 28(3): 569-582.
- Wheeler S.M., 2000. Planning for metropolitan sustainability. *Journal of Planning Education and Research*, 20:133-45.
- White R., 2006. Pattern based map comparisons. *Journal of Geographical systems*, 8: 145-164.
- White R. and G. Engelen, 1997. Cellular automata as the basis of integrated dynamic regional modeling. *Environment and Planning B*, 24:235-246.
- White R. and G. Engelen, 2000. High-resolution integrated modeling of the spatial dynamics of urban and regional systems. *Computer, Environment and Urban Systems*, 24, 383-400.
- White R., Uljee I. and G. Engelen, 2012. Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *International Journal of Geographical Information Science*, 26(7):1251-1280.

- White R., Engelen G., and I. Uljee, 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*, 24 (3), 323-343.
- Whitehead C.J., 1968. *Country Residential growth in Calgary region: a study of exurbanization*. Master dissertation. The University of British Columbia.
https://circle.ubc.ca/bitstream/handle/2429/36135/UBC_1968_A8%20W55.pdf?sequence=1
- Wijesekara G.N., A. Gupta, Valeo C., Hasbani J.-G., Qiao Y., Delaney P. and D.J. Marceau, 2012. Assessing the impact of future land-use changes on hydrological processes in the Elbow River watershed in southern Alberta, Canada. *Journal of hydrology*, 412-413: 220-232.
- Wikipedia, 2012. <http://en.wikipedia.org/wiki/Calgary>.
- Wolfram S., 1984. Computation theory of cellular automata. *Communications in Mathematical Physics*, 96: 15-57.
- Wu F., 1998. SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rule. *International Journal of Geographical Information Science*, 12: 63-82.
- Wu F. 2002. An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B*, 25(5):731-752.
- Wu D., Liu J., Zhang, G., Ding W., Wang W. and R.Wang, 2009. Incorporating spatial autocorrelation into cellular automata model: an application to the dynamics of Chinese tamarisk. *Ecological Modelling*, 220(24): 3490-3498.

- Wu N. and E.A. Silva, 2010. Artificial intelligence solutions for urban land dynamics: a review. *Journal of Planning Literature*, 24(3): 246-265.
- Xie Y., 1996. A generalized model for cellular urban dynamics. *Geographical Analysis*, 28: 350-373.
- Xie Y., Fang C., Lin G.C., Gong H. and B. Qiao, 2007. Tempo-Spatial Patterns of Land Use Changes and Urban Development in Globalizing China: A Study of Beijing. *Sensors*, 7(11):2881-2906.
- Yang Q., Li X. and X. Shi, 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers and Geosciences*, 34: 592-602.
- Yang Q. and X. Li, 2006. Mining transition rules for Geo-simulation using rough sets. *Acta Geographica Sinica*, 61(8), 882-894.
- Yang X. and C.P. Lo, 2003. Modeling urban growth and landscape change for Atlanta metropolitan region. *International Journal of Geographical Information Science*, 17(5): 463-488.
- Zhao G., 2011. The effect of land use categories resolution on geography cellular automata model. *Proceedings of the International Conference on Multimedia Technology (ICMT)*, Hangzhou, July 26-28.
- Zhao F.Y., Jia Y.F., Shao, B. and Y.H. Zhu, 2012. Forecast of Dynamic Change of Land Use Based on Cellular Automata. *Asian Agricultural Research*, 4(3): 19-22.
- Zhang Q., Ban Y., Liu J. and Y. Hu, 2011. Simulation and analysis of urban growth scenarios for the Greater Shanghai Area, China. *Computers, Environment and Urban Systems*, 35:126-139.

Appendix

1. Rules generated using RST for the transition from forest to built-up:

Where:

‘a’ represents the factor ‘Distance to road’;

‘b’ represents the factor ‘Distance to river’;

‘c’ represents the factor ‘Number of built-up cells in NH0’;

‘d’ represents the factor ‘Number of vegetation cells in NH1’;

‘e’ represents the factor ‘Number of built-up cells in NH1’;

‘f’ represents the factor ‘Number of built-up cells in NH2’;

‘g’ represents the factor ‘Number of Tsuu T’ina Nation land cells in NH2’;

‘V1’ represents ‘percentage of cells updated by each transition rule’;

‘V2’ represents ‘percentage of change to built-up by the transition rule’;

F(5) represents ‘final state is built-up’;

F(4) represents ‘final state is vegetation’.

The numbers 1,2,3,4 in parentheses associated to a, b, c, d, e, f, and g represent to which discretization categories the factor belongs.

Rule	V1(%)	V2(%)
a(2)AND b(2)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.41	2.48
a(1)AND b(3)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.30	1.83
a(1)AND b(2)AND c(4)AND d(1)AND e(3)AND f(4)AND g(1) => F(5)	0.24	1.47
a(2)AND b(1)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.18	1.10
a(2)AND b(1)AND c(4)AND d(1)AND e(3)AND f(3)AND g(1) => F(5)	0.18	1.10
a(2)AND b(3)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.15	0.92
a(3)AND b(2)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.14	0.83
a(3)AND b(1)AND c(4)AND d(1)AND e(4)AND f(4)AND g(1) => F(5)	0.12	0.73
a(1)AND b(2)AND c(4)AND d(1)AND e(3)AND f(3)AND g(2) => F(5)	0.12	0.73
a(2)AND b(2)AND c(1)AND d(2)AND e(2)AND f(3)AND g(2) => F(5)	0.12	0.73
a(1)AND b(2)AND c(4)AND d(1)AND e(4)AND f(3)AND g(2) => F(5)	0.11	0.64
a(2)AND b(1)AND c(4)AND d(2)AND e(3)AND f(3)AND g(1) => F(5)	0.11	0.64
a(1)AND b(2)AND c(4)AND d(2)AND e(4)AND f(3)AND g(1) => F(5)	0.11	0.64
a(1)AND b(2)AND c(4)AND d(2)AND e(4)AND f(4)AND g(1) => F(5)	0.09	0.55
a(2)AND b(3)AND c(4)AND d(2)AND e(3)AND f(4)AND g(1) => F(5)	0.09	0.55
a(1)AND b(2)AND c(4)AND d(1)AND e(4)AND f(3)AND g(1) => F(5)	0.08	0.46
a(1)AND b(2)AND c(4)AND d(1)AND e(3)AND f(4)AND g(2) => F(5)	0.08	0.46
a(1)AND b(1)AND c(4)AND d(1)AND e(4)AND f(3)AND g(2) => F(5)	0.08	0.46
a(1)AND b(1)AND c(4)AND d(1)AND e(2)AND f(3)AND g(2) => F(5)	0.08	0.46
a(1)AND b(2)AND c(4)AND d(1)AND e(2)AND f(3)AND g(2) => F(5)	0.08	0.46
a(1)AND b(3)AND c(1)AND d(3)AND e(1)AND f(2)AND g(1) => F(4) XOR F(5)	4.67	3.76
a(2)AND b(1)AND c(3)AND d(2)AND e(3)AND f(3)AND g(1) => F(4) XOR F(5)	0.72	3.49
a(1)AND b(2)AND c(4)AND d(2)AND e(3)AND f(3)AND g(2) => F(4) XOR F(5)	0.86	3.39
a(2)AND b(1)AND c(3)AND d(2)AND e(3)AND f(3)AND g(2) => F(4) XOR F(5)	0.50	2.57
a(1)AND b(1)AND c(2)AND d(2)AND e(2)AND f(3)AND g(2) => F(4) XOR F(5)	1.12	2.39

a(1)AND b(1)AND c(4)AND d(1)AND e(3)AND f(3)AND g(2) => F(4) XOR F(5)	0.41	2.29
a(1)AND b(3)AND c(1)AND d(3)AND e(2)AND f(2)AND g(1) => F(4) XOR F(5)	1.25	2.20
a(2)AND b(1)AND c(2)AND d(2)AND e(2)AND f(2)AND g(1) => F(4) XOR F(5)	2.26	2.11
a(1)AND b(2)AND c(2)AND d(2)AND e(2)AND f(3)AND g(2) => F(4) XOR F(5)	0.39	1.56
a(3)AND b(1)AND c(1)AND d(2)AND e(2)AND f(2)AND g(1) => F(4) XOR F(5)	0.24	1.38
a(1)AND b(3)AND c(2)AND d(3)AND e(2)AND f(2)AND g(1) => F(4) XOR F(5)	0.78	1.28
a(1)AND b(2)AND c(2)AND d(3)AND e(2)AND f(3)AND g(1) => F(4) XOR F(5)	1.13	1.28
a(1)AND b(2)AND c(3)AND d(2)AND e(2)AND f(3)AND g(2) => F(4) XOR F(5)	0.32	1.28
a(1)AND b(2)AND c(2)AND d(3)AND e(2)AND f(3)AND g(2) => F(4) XOR F(5)	0.30	1.28
a(1)AND b(1)AND c(3)AND d(2)AND e(3)AND f(3)AND g(2) => F(4) XOR F(5)	0.24	1.19
a(1)AND b(1)AND c(2)AND d(2)AND e(2)AND f(3)AND g(1) => F(4) XOR F(5)	1.04	1.10
a(1)AND b(2)AND c(3)AND d(2)AND e(3)AND f(3)AND g(2) => F(4) XOR F(5)	0.39	1.10
a(1)AND b(2)AND c(4)AND d(2)AND e(2)AND f(3)AND g(2) => F(4) XOR F(5)	0.27	1.10
a(1)AND b(2)AND c(2)AND d(2)AND e(2)AND f(2)AND g(2) => F(4) XOR F(5)	0.98	1.10