2007

Land-Cover and Land-Use Study Using Genetic Algorithms, Petri Nets, and Cellular Automata

Fei Wang
Louisiana State University and Agricultural and Mechanical College, fwang3@lsu.edu

LAND-COVER AND LAND-USE STUDY USING GENETIC ALGORITHMS, PETRI NETS, AND CELLULAR AUTOMATA

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geography and Anthropology

by
Fei Wang
B.S., Wuhan Teacher's College, 1984
M.S., Beijing Normal University, 1995
M.S., Southern University, 2002
December, 2007

ACKNOWLEDGEMENTS

During my years in Baton Rouge, Louisiana, I have been blessed with guidance, help, and encouragement from more people than I could possibly cite in this acknowledgement. I deeply believe that I am indebted to those who have guided, helped, and encouraged me.

Primary thanks must go to my major advisor Professor Nina Lam for her guidance during last four and half years. Her suggestions have greatly kindled my interest in artificial intelligence-based spatiotemporal analysis and modeling. I am thankful to my committee members – Professor Kam-Biu Liu, Mr. DeWitt Braud, Professor Robert V. Rohli, and Professor Max Z. Conrad – for their invaluable support. They not only introduce knowledge but also teach the way of thinking.

Special thanks to Professor Qingxin Lin at Wetland Biogeochemistry Institute at Louisiana State University and Professor Exyie C. Ryder at the Department of Science and Mathematics Education at Southern University for all the help they have provided. Special thanks also to Professor Tony Lewis, Professor Craig E. Colten, Professor John Pine, Professor Andrew Curtis, and Mr. Farrell Jones at Department of Geography and Anthropology at Louisiana State University for their help. Without their help, I would not be able to finish this study.

Thanks to all faculty and staff at the Department of Geography and Anthropology at Louisiana State University, all faculty and staff at the Department of Science and Mathematic Education at Southern University, and all faculty and staff at the Department of Social Sciences at Louisiana Tech University.

Last but not least, I would like to thank my parents and parents-in-law, my wife and daughter, my brother, my sister and her family, and my brother-in-law and his family for their love, understanding, encouragement, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Recent research techniques, such as genetic algorithm (GA), Petri net (PN), and cellular automata (CA) have been applied in a number of studies. However, their capability and performance in land-cover land-use (LCLU) classification, change detection, and predictive modeling have not been well understood. This study seeks to address the following questions: 1) How do genetic parameters impact the accuracy of GA-based LCLU classification; 2) How do image parameters impact the accuracy of GA-based LCLU classification; 3) Is GA-based LCLU classification more accurate than the maximum likelihood classifier (MLC), iterative self-organizing data analysis technique (ISODATA), and the hybrid approach; 4) How do genetic parameters impact Petri Net-based LCLU change detection; and 5) How do cellular automata components impact the accuracy of LCLU predictive modeling.

The study area, namely the Tickfaw River watershed (711mi^2), is located in southeast Louisiana and southwest Mississippi. The major datasets include time-series Landsat TM / ETM images and Digital Orthophoto Quarter Quadrangles (DOQQ's). LCLU classification was conducted by using the GA, MLC, ISODATA, and Hybrid approach. The LCLU change was modeled by using genetic PN-based process mining technique. The process models were interpreted and input to a CA for predicting future LCLU.

The major findings include: 1) GA-based LCLU classification is more accurate than the traditional approaches; 2) When genetic parameters, image parameters, or CA components are configured improperly, the accuracy of LCLU classification, the coverage of LCLU change process model, and/or the accuracy of LCLU predictive modeling will be low; 3) For GA-based LCLU classification, the recommended configuration of genetic / image parameters is generation 2000-5000, population 1000, crossover rate 69%-99%, mutation rate 0.1%-0.5%, generation gap 25%-50%, data layers 16-20, training / testing data size 10000-20000 / 5000-10000, and spatial

resolution 30m-60m; 4) For genetic Petri nets-based LCLU change detection, the recommended configuration of genetic parameters is generation 500, population 300, crossover rate 59%, mutation rate 5%, and elitism rate 4%; and 5) For CA-based LCLU predictive modeling, the recommended configuration of CA components is space 6025 * 12993, state 2, von Neumann neighborhood 3 * 3, time step 2-3 years, and optimized transition rules.

CHAPTER 1 INTRODUCTION

1.1. Background

1.1.1. Remote Sensing and GIS

Remote sensing is not only the technology and application of installing sensors on aircrafts or satellites for acquiring objects' data without physical contact, but also the science and art of extracting, analyzing, and discovering information and knowledge from objects' data (Campbell, 1996; Tso and Mather, 2001). The objects' data are represented by electromagnetic energy reflected from or emitted by the Earth's surface (Tso and Mather, 2001). The energy come from the and it is associated with a wide range wavelengths of electromagnetic spectrum (Lo and Yeung, 2007; Tso and Mather, 2001). The important characteristics of remote sensing are that it is the technique of data collection and processing, the science of data analysis and pattern recognition, and the art of information representation. Since the 1970s, remote sensing has increasingly become one of the greatest approaches that scientists use to collect data, recognize patterns, detect changes, and develop models throughout the entire world at different spatiotemporal scales. Both satellite images and aerial photographs have become one of major data sources for designing and building geodatabases (Campbell, 1996).

During the early stage in remote sensing development, GIS was conceptualized as a class of systems. Unfortunately, they did not become operational until 1971. From the early 1980s to the mid-1990s, GIS became mature in terms of theories, methodologies, and applications. Since the mid-1990s, the development of GIS has been greatly accelerated by the rapid growth of computer technology. Now, GIS commonly includes at least four aspects: science, system, study, and service (DiBiase *et al.*, 2006). It is not only a set of applications and studies relating to the real world problems, or a special type of information systems supporting the collection, storing, management, processing, analysis, modeling, and display of geospatial data for solving the real world problems (Lo and Yeung, 2007), but also a science including multidisciplinary research that address the fundamental theoretic, methodological, and practical problems (Goodchild, 1992).

Despite both remote sensing and GIS having achieved obvious success, there are a variety of problems affecting their implementation, including data accessibility, user diversity, technology compatibility, and application variety. Integration of remote sensing and GIS has been recognized as one of the better ways to address those problems (Hinton, 1996; Lo and Yeung, 2007). An important area of integrating remote sensing and GIS lies in introducing vector information into the procedure of image classification (Ehlers *et al.*, 1989). Essentially, it is a procedure to identify pixels falling on or in a particular vector feature (*e.g.*, line or polygon) so that more geospatial information can be provided (Lo and Yeung, 2007).

1.1.2. Land-Cover and Land-Use Study

The interaction between human society and the natural environment is complex, and land-cover and land-use (LCLU) has played a critical role in the course of such interaction. Since early 20th century, human activities and natural forces have significantly affected biogeochemical processes at various scales substantially. Global warming, natural resource shortage, environmental hazard increase, and biodiversity decrease are not scientific hypotheses anymore. They have become the real urgent crises (Dolman and Verhagen, 2003). LCLU is recognized as a linkage between human activities and natural processes, and LCLU change is one of the major drivers of global environmental change (Read and Lam, 2002). LCLU study plays an important role in the process of improving the nature-society relationship. It includes three tasks: classification, change detection, and predictive modeling. The importance, complexity, and uncertainty of LCLU make it an ideal illustration of spatiotemporal study. LCLU classification is a procedure to identify a land feature in a particular area at a particular time; LCLU changes are usually measured as the differences in the states of land features observed at different times (Singh, 1989); and LCLU predictive modeling is the procedure to model future LCLU change based on time series dataset. To detect and predict LCLU change, an accurate LCLU classification should be carried out.

1.1.3. Traditional Land-Cover and Land-Use Study Approaches

It is commonly believed that LCLU study includes classification, change detection, and predictive modeling. LCLU classification is a process of extracting land feature information using remotely-sensed multi-spectral images. Traditionally, there are three classification approaches including supervised, unsupervised, and hybrid (Campbell, 1996; Lo and Yeung, 2007). These traditional approaches, fully or partially based on statistics, have been successfully used in many LCLU classification studies. Unfortunately, as human abilities of collecting and storing geospatial data increase, the availability of plentiful geospatial data sources greatly challenge human capabilities of extracting information and discovering knowledge about land features (Huang, 1996; Huang and Jenson, 1997). The maximum likelihood classifier (MLC), one of the supervised classification techniques, assumes that the data for each class are normally distributed (Huang, 1996; Huang and Jenson, 1997). The iterative self-organizing data analysis technique (ISODATA), one of the unsupervised classification approaches, requires subsequent identification of LCLU features. Although the hybrid approach, based on both MLC and ISODATA, can partially solve the problem in either supervised or unsupervised classification, essentially, the problem still exists.

Another common application of remote sensing technology is LCLU change detection using time series LCLU maps (Singh, 1989). Since the 1980s, a number of change detection techniques have been developed, and due to the importance of LCLU change detection, many new techniques have been constantly proposed. Based on the studies done by several scholars

(Chan *et al*., 2001; Civco *et al*., 2002; Coppin *et al*., 2004; Lu *et al*., 2004a; Mas, 1999; and Singh, 1989), most LCLU change detection techniques could be categorized into the following groups: post-classification comparison, composite analysis, univariate image differencing, image ratioing, bi-temporal linear data transformation, change vector analysis, image regression, multi-temporal spectral mixture analysis, and multidimensional temporal feature space analysis. Although most of them have been efficiently used to detect LCLU change, such as the type and location of land feature, and rate of land features change, they are not efficient in discovering change processes.

Studies on LCLU classification and change detection are numerous. However, there are very few studies on LCLU predictive modeling (Li and Yeh, 2002 and 2004). LCLU predictive modeling is the procedure of forecasting future LCLU based on the existing observation data. The major traditional approaches include empirical-statistical models and stochastic models (Lambin, 2004). Empirical-statistical models (*e.g*., multiple linear regression, logistic regression model) are based on the unrealistic assumption of linear relationships between predictor variables and dependent variable (Brown *et al*., 2004; Lambin, 2004). Stochastic models (*e.g*., Markov chains) rely on the assumption of the stationary of the transition matrix, and the model has little explanatory power (Lambin, 2004).

1.1.4. Non-Traditional Land-Cover and Land-Use Study Approaches

Although the big debate about the role of systemic and quantitative methods in geographic study area has been over for half a century, the study of mathematic models and quantitative methods in geography has been enhanced tremendously (Johnston, 1997; Legates *et al*., 2003), our ability in geospatial information extraction and knowledge discovery had been limited until the mid-1990s. Since the mid-1990s, and because of the rapid development of computer technology and increasing need solving geographic problems, many new techniques have been introduced in LCLU studies, such as the machine learning-based approach and geospatial knowledge discovery (or geospatial data mining).

Since the early 1990s, the focus of LCLU classification has shifted from spectral analysis to contextual, spatial syntactic analysis, and to knowledge-based interpretation (Argialas and Harlow, 1990; Huang and Jenson, 1997). Among different knowledge-based LCLU classification techniques such as decision tree, fuzzy logic, association rule, support vector machine, and neural networks, genetic algorithm (GA) is one of the most promising techniques intensively applied to LCLU classification (Perkins *et al*., 2000; Stathakis and Vasilakos, 2006; Tso and Mather, 1999). These approaches have been proved effective in improving classification accuracy (Lu and Weng, 2007). A critical step is to develop the rules that can be used in an expert system or a knowledge-based classifier. This approach has now become increasingly attractive because of its capability of accommodating multiple sources of data (Lu and Weng, 2007). Usually, there are three ways to build rules for image classification: 1) to elicit and refine explicit knowledge and rules from experts, 2) to extract implicit knowledge and rules using cognitive methods, and 3) to generate empirical knowledge and rules from observed data using automatic induction methods (Hodgson *et al*., 2003). Both GIS and remote sensing play an important role in developing knowledge-based classification approaches because of its capability of managing different sources of data and spatial modeling.

The relationship between LCLU classification and change detection is very close, and some research even mix both together and regards classification as a portion of LCLU change detection (Coppin *et al*., 2004; Lu *et al*., 2004a; Lu and Weng, 2007); however, they are different in terms of purposes, tasks, and procedures. The task of classification is to identify the pattern of LCLU at a particular time, and the task of change detection is to identify the difference of LCLU during a particular time period. Unlike LCLU classification, the applications of artificial intelligence in LCLU change detection have seldom been reported. It is commonly recognized that a systematic LCLU change detection should answer the following questions: 1) What types of land features change? 2) When do land features change? 3) Where do land features change? 4) How do land features change? and 5) Why do land features change (Coppin *et al*., 2004; Lu *et al*., 2004a; Miller *et al*., 1998; Petit *et al*., 2001)? 6) What is the accuracy of LCLU change detection? Most existing LCLU change detection studies have done a great job in terms of providing information about the LCLU differences between two or more end points. This information can be used to answer the first three questions, but it is not enough for answering the last two questions. To answer the last two questions, the processes of LCLU change should be discovered (Miller *et al*., 1998; Petit *et al*., 2001).

The dynamics of LCLU change processes can be investigated through a temporal series of remote sensing data and by analyzing the relationship among various LCLU patterns derived from different years (Turner, 1987; Singh, 1989; Hall *et al*., 1991; Coppin and Bauer, 1996; Lambin, 1996; Mertens and Lambin, 2000; Petit *et al*., 2001). Multi-temporal analysis may be easy to understand but hard to implement, especially when the amount of land feature or observation year is large (Miller *et al*., 1998). Markov chains may be used efficiently for quantifying LCLU processes, but the resultant process information is implicitly hidden in the mathematical models (Petit *et al*., 2001). LCLU change may improve or damage the relationship between human and nature. To prevent harmful LCLU change or decrease its impact, it is necessary to know the future of LCLU change. From the day von Neumann and Ulam proposed the concept of cellular automata (CA) (Neumann, 1966), to Wolfram's recent book *A New Kind of Science* (Wolfram, 2002), the simple structure of cellular automata has attracted researchers from various disciplines. Geographers have used cellular automata in LCLU predictive modeling for many years. So far, a number of CA-based LCLU predictive models have been developed specially for LCLU change studies, such as the research institute for knowledge systems (RIKS) model (Engelen *et al*., 1997), slope, land-use, exclusion, urban extent, transportation, and hill shade (SLEUTH) model (Clarke *et al*., 1997), Fuzzy-CA (Wu, 1998b), ANN-CA (Li and Yeh, 2002), multicriteria evaluation-cellular automata (MCE-CA) (Wu and Webster, 1998), statistic CA (Sui and Zeng, 2001), and stochastic CA (Wu, 2002).

Cellular automata are decentralized computing models that are composed of five elements: 1) Space – the cell space is composed of individual cell, and most cellular automata adopt regular grids to represent such a space; 2) States – each cell may represent any variable, e.g., the various types of land features; 3) Time steps – a cellular automaton will evolve at a sequence of discrete time steps, and at each step the cells will be updated simultaneously based on transition rules; 4) Neighborhood – each cell has neighbors around it, and its state will be affected by its neighbors; and 5) Transition rules – they are key components of cellular automata, and a transition rule is normally specified to change the states of cell (Neumann, 1966; Wolfram, 2002). Transition rule is the critical component of cellular automata. The difficulty of designing cellular automata’s transition rules to solve a particular problem has severely limited their applications (Kanoh and Wu, 2003). Although there are virtually unlimited amount of ways to define transition rules, the transition rules of traditional cellular automata are usually represented by mathematical equations (Li and Yeh, 2004). Defined by mathematical equations, the transition rules in many cases are not explicit, hard to understand, and difficult to implement (Li and Yeh, 2004). Data mining or machine learning may be a good approach to develop a set of explicit transition rules for a CA.

1.2. Problems Statement

During the last decade, genetic algorithms (GAs) have been introduced into numerous LCLU studies, and most of them refer to LCLU classification. Genetic algorithms were proposed as a class of random search and global optimization techniques, but they do not automatically inherit the characteristics of global optimization from either Darwin's evolutionary theory or modern genetics. In spite of great achievement, GAs-based approaches still suffer from a number of deficiencies. The most harmful one is that any improper setting of control parameters could cause the population to lose its diversity, be dominated by so-called elites, and prematurely converge to local optimization (Wu and Cao, 1997; Rocha and Neves, 1999). Theoretically, the improper configuration of genetic parameters can lead to Genetic algorithms performing poorly (*e.g.*, premature convergence / local optimization) and impact the accuracy of LCLU classification, but the existing GA-based LCLU classifications have seldom investigated such GA's limitation.

Moreover, although it is commonly recognized that a systematic and accurate LCLU change detection should provide the information of change processes, only a few studies have done that. The existing process-oriented LCLU change detection studies still have some limitations: 1) only a few land features or classes were involved (Miller *et al*., 1998); 2) information on the resultant processes were not explicit and difficult to interoperate (Petit *et al*., 2001); and 3) the resultant process models were not optimized in terms of representing more LCLU change process instances. There are many LCLU classifications using genetic algorithms, but process-oriented LCLU change detections using genetic algorithms have not been found so far. The major reason for using genetic algorithms in this procedure is to optimize the change process model, so that the resultant process models can best represent real LCLU change processes. As it has been pointed out in the previous section, the transition rule is a critical component of a cellular automaton (CA), and data mining or machine learning technique may be a better approach to discover transition rules for a cellular automaton. Decision trees have been used to discover transition rules for a geographical cellular automaton (Li and Yeh, 2004). A decision tree can learn based on the given data set, but it cannot improve the learning by increasing search space and optimizing resultant knowledge. This dissertation seeks to address the following five groups of problems, and each group of problems includes several subgroup problems:

- How do genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generation gap) impact the accuracy of GA-based LCLU classification?
- How do image parameters (such as spatial resolution, training data size, different indexing data, and data combination) impact the accuracy of GA-based LCLU classification?
- Is GA-based LCLU classification more accurate than maximum likelihood classifier (MLC) based supervised classification, iterative self-organizing data analysis technique (ISODATA) based unsupervised classification, or the hybrid of both MLC and ISODATA?
- How do genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generation gap) impact the accuracy of process-oriented change detection?
- How do cellular automata components (such as space, states, neighborhood, time steps, and transition rules) impact the accuracy of LCLU predictive modeling?

1.3. Research Objectives

The major goal of the dissertation is to develop a systematic and better understanding on genetic algorithms, Petri nets (PNs), and cellular automata-based LCLU study. There are five objectives in this study. Each objective was related to a particular research problem. Specifically, the research objectives include:

- Examine the relationships between genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generations gap) and GA-based LCLU classification.
- Examine the relationships between image parameters (such as spatial resolution, training / testing data size, different indexing data, and data combination) and GA-based LCLU classification.
- Compare GA-based LCLU classification, MLC-based supervised classification, ISODATA-based unsupervised classification, and the hybrid of both MLC and ISODATA.
- Examine the relationships between genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generations gap) and process-oriented LCLU change detection.
- Examine the relationship between cellular automata components (such as space, states, time steps, neighborhood, and transition rules) and predictive modeling. The major concern is to test the performance of the transition rules that were derived by PN, optimized by GA, and implemented by CA.

1.4. Research Hypotheses

Based on the objectives of the research, the primary hypotheses of this study were divided into five sets also: 1) genetic parameters impact classification, 2) image parameters impact classification, 3) GA-based approaches are better than traditional approaches, 4) genetic parameters impact process-oriented change detection, and 5) CA components impact predictive modeling. Each group of hypotheses was related to a particular research problem and objective. Each hypothesis contains hull hypothesis and alternate hypothesis. In order to briefly and clearly describe the hypothesis, only alternate hypothesis is listed.

- For examining the impact of genetic parameters on the accuracy of GA-based LCLU classification, hypothesis #1 is as follows:

  Ha: Different number of generations, population size, crossover rate, mutation rate, and generations gap can increase / decrease the accuracy of GA-based LCLU classification.

- For examining the impact of image parameters on the accuracy of GA-based LCLU classification, hypothesis #2 is as follows:

  Ha: Different spatial resolution, training / testing data size, data layers, and data layer combinations can increase / decrease the accuracy of GA-based LCLU classification.

- For examining whether GAs-based LCLU classification is more accurate than traditional LCLU classification approaches, hypothesis #3 is as follows:

  Ha: GAs-based LCLU classification is more accurate than MLC-based supervised classification, ISODATA-based unsupervised classification, and the hybrid of MLC and ISODATA.

- For examining the impact of genetic parameters derived from Petri nets on process-oriented LCLU change detection, hypothesis #4 is as follows:

  Ha: Different number of generations, population size, crossover rate, mutation rate, or generations gap can increase / decrease the accuracy of genetic PN-based LCLU change process detection.

- For examining the impact of cellular automata components (such as space, states, time steps, neighborhood, and transition rules) increase / decrease the accuracy of LCLU predictive modeling, hypothesis #5 is as follows:

  Ha: Different space size, state size, time steps, neighborhood type and size, or transition rules can increase / decrease the accuracy of predictive modeling.

1.5. Expected Significance

The research is among the first to: 1) examine the relationships among the genetic parameters, image parameters, and LCLU classification; 2) determine the processes of LCLU change using genetic Petri nets (PN); and 3) predict future LCLU using genetic PN-based transition rules in a geographic cellular automaton. The expected significance of the research includes the following three aspects:

- **Theoretical Significance**

Although LCLU study has a long-term history in the GIS and remote sensing community, due to the spatial, temporal, and phenomena complexity, the processes of LCLU change are still not well-understood. A set of explicit and systematic spatio-temporal knowledge will definitely improve human understanding of LCLU change. Also, the study will improve our understanding of genetic algorithms. GAs are inspired by the theory of genetics and natural selection; they do not automatically inherit the characteristics of global optimization from both. It is important to recognize that improper parameter configuration can make GAs perform poorly.

- **Methodological Significance**

One of the most important tasks in LCLU change detection is to identify the change processes. Previous research has indicated that there is no standard approach to perform process-oriented LCLU change detection in terms of efficiency, capability, and explicit results. The integration of GA-based Petri nets and GIS provides a very good way to perform such study. Also, the dissertation research proposes a new approach to develop transition rules for geographic cellular automata. The transition rules developed from GA-based Petri nets are not only explicit but also optimized in terms of correctly representing approximating change processes. This approach can have applications in geography, environmental science, ecology, economics, and politics.

- **Practical Significance**

The study area, namely Tickfaw watershed, has experienced tremendous urban growth and forest fragmentation during the last twenty years. The human-nature relationship has become one of most important problems that impact its regional development. This research provides both current and future LCLU maps that can be used by the local government for analyzing environmental quality, managing natural resources, and making the regional sustainable development planning.

1.6. Rationale of Dissertation Research

- **Why Land-Cover and Land-Use Study?**

As mentioned earlier in this chapter, LCLU is one of most important representation of human-nature relationships. The importance, complexity, and uncertainty of LCLU change needs further research. Also, the processes of LCLU change are always impacted by human activity and natural processes. It is always of high theoretical and practical priority to get a better understanding of interaction between human and nature.

- **Why Genetic Algorithms, Petri Nets, and Cellular Automata?**

LCLU classification, change detection, and predictive modeling are different decision-making processes. Optimization is always important for making a decision. Genetic algorithms were proposed as a random search and global optimization technique (Goldberg, 1989; Holland, 1975). A process-oriented LCLU change detection focus on the processes determination, and Petri nets is a mathematical and graphic language especially designed for modeling the structure and behavior of a system. LCLU change is a complexity issue, and traditional approaches (such as regression or Markov chain) have limitations in representing such complexity (Petit *et al*., 2001; Lambin, 2004). Cellular automata are flexible modeling systems, and they have long term successful history in spatiotemporal modeling (Li and Yeh, 2004).

- **Why the Particular Study Area and Data Sources?**

Tickfaw watershed, located in southeastern Louisiana, is one of the watersheds in the upper Lake Pontchartrain basin. It includes many different LCLU features. Since the middle 1980s, it has experienced great change in terms of urban growth and forest segmentation. The major data sources include Landsat Thematic Mapper / Enhanced Thematic Mapper (TM / ETM) images, digital orthophoto quarter quadrangles (DOQQs), digital elevation model (DEM), and soil data. Those data have very high availability and accessibility.

- **Scope of Dissertation Research**

A LCLU study involves many theoretical, methodological, and practical problems. Although the generality, completeness, and significance of the research is recognized, due to resource limitation, time limitation, and capability limitation, this dissertation research only focuses: 1) spatially on the Tickfaw River watershed, 2) temporally on the period of 1986-2015, 3) theoretically on LCLU classification, change detection, and predictive modeling, and 4) methodologically on genetic algorithms, Petri nets, and cellular automata.

1.7. Chapter Organization

The dissertation comprises nine chapters. The current chapter gives an overview of the research. The study background provides basic information about current content and progress in each research sub-area. It indicates the importance of LCLU study, the problems faced by traditional approaches, and the problems faced by non-traditional approaches. The study seeks to address five problems related to LCLU classification, change detection, and predictive modeling using GAs, PNs, and CA-based approaches. According to those problems, the study objectives and the corresponding research hypotheses are laid out. The study will examine the relationships among genetic parameters, image parameters, PNs, CA components, and performance of LCLU study. The expected significance includes theory, methodology, and practice aspects.

The second chapter reviews previous studies relating to this dissertation research. It includes the development of remote sensing and GIS, development of LCLU study, and development of GAs, PNs, and CA-based approaches. For each section in this chapter, the current research progress, advances, and limitations are discussed.

The third chapter illustrates the study area, data sources, algorithms, implementation, experimental design, and research procedures. Data preprocessing is a critical step for this study. After geometric and radiometric correction, the indexing data and transformed data were created. All algorithms were integrated with different GIS and remote sensing software.

From Chapters 4 to 8, different experiments were conducted to test the hypotheses, and the results are presented in individual chapters. The impacts of both GA parameters and image parameters on the accuracy of LCLU classification, the impacts of GA parameters on the accuracy of process-oriented change detection, and the impacts of cellular automata components on the accuracy of LCLU predictive modeling based on GA-Petri nets derived transition rules are discussed respectively.

The last chapter summarizes the major findings from different chapters, the contribution of the dissertation research, and provides some recommendations for future studies.

CHAPTER 2 LITERATURE REVIEW

2.1. Land-Cover and Land-Use Classification

2.1.1. Concept of Land-Cover and Land-Use and Classification

Land cover refers to the natural materials existing on the earth's surface, land use describes the way humans use the land surface, and both of them closely links human activities to natural process (Read and Lam, 2002). The importance, complexity, and uncertainty of land cover and land use make it an ideal illustration of geographic research. LCLU study – including classification, change detection, and predictive modeling – plays a crucial role in the process of improving the social-natural relationship (Dolman and Verhagen, 2003). The goal of LCLU classification is to identify LCLU features within a particular space at a particular time or period. It is the foundation of change detection and predictive modeling. Based on the survey done by Lu and Weng (2007), a systematic LCLU classification should include at least the following seven steps: 1) acquire or collect a suitable remotely-sensed data set and ancillary data set, 2) select a suitable classification system, 3) perform necessary data preprocessing, 4) take good training samples, 4) develop classification knowledge, 5) extract LCLU features information and perform classification, 6) perform post-classification processing and estimate the accuracy.

2.1.2. Approaches to Land-Cover and Land-Use Classification

Various LCLU classification techniques have been developed, and many have been summarized and reviewed (Coppin *et al.*, 2004; Lu *et al.*, 2004a; Lu and Weng, 2007). Several criteria can be used for categorizing LCLU classification techniques: 1) pixels, 2) training samples, 3) information types, 4) output, 5) parameters, 6) artificial intelligence, 7) knowledge, and 8) combination (Table 2.1) (Lu and Weng, 2007).

The traditional LCLU classification techniques are based on statistics (such as minimum distance to means classifier, maximum likelihood classifier, and K-mean clustering) have

Table 2.1 A taxonomy of land-cover and land-use classification techniques (Source: modified from Lu and Weng, 2007)

Criteria	Categories	Examples	References
Pixels	Subpixel-based classification	Subpixel classifiers	Huguenin *et al*., 1997
		Linear regression	Settle and Campbell, 1998
	Pixel-based classification	Most of classifiers	Lu and Weng, 2007
		Multiple-forward-model	Peddle *et al*., 2004
	Object-oriented classification	eCognition	Gao *et al*., 2006
		Object-oriented approach	Geneletti and Gorte, 2003
	Field-based classification	Parcel-based approaches	Dean and Smith, 2003
		Map-guided classification	Chalifoux *et al*., 1998
Training samples	Supervised classification	Maximum likelihood	Jensen, 1996
		Parallepiped classifier	Lo & Yeung, 2007
	Unsupervised classification	ISODATA	Jensen, 1996
		K-mean clustering	Stuckens *et al*., 2000
	Hybrid classification	Hybrid classifiers	Lo and Yeung, 2007
		Hybrid approach	Stuckens *et al*., 2000
Parameters	Parametric classification	Maximum likelihood,	Jensen, 1996
		Discriminant analysis	Lo and Yeung, 2007
	Non-parametric classification	Neural network	Chen, *et al*., 1995
		Neural network	Foody *et al*., 1995
		Neural network	Verbeke et al., 2004
		Decision tree classifier	DeFries *et al*., 1998
Information	Spectral classification	Maximum likelihood	Jensen, 1996
		Hybrid classifier	Lo and Yeung, 2007
	Contextual classification	Freq-contextual classifier	Kartikeyan *et al*., 1994
		Spatio-spectral classifier	Schowengerdt, 1996
	Spectral-spatial classification	ECHO	Lu *et al*., 2004a and 2004b
		Hybrid approach	Stuckens *et al*., 2000
	Spectral-temporal classification	Multi-temporal approach	Wolter *et al*., 1995
		Multi-temporal approach	Tottrup, 2004
Output definiteness	Hard classification	Maximum likelihood	Jensen, 1996
		Decision tree	DeFries *et al*., 1998
	Soft classification	Fuzzy logic classifiers	Shalan *et al*., 2003
		Visual fuzzy classifier	Lucieer and Kraak, 2004

limitations when dealing with spatial data that have the properties of inaccuracy, multiple scale, and interdependency. Since the early 1990s, the focus of LCLU classification techniques has been shifted from purely spectral analysis to contextual, spatial syntactic analysis, and knowledge-based interpretation (Argialas and Harlow, 1990; Srinivasan and Richards, 1990; Kontoes *et al*., 1993; Huang and Jenson, 1997; Legates *et al*., 2003). For the knowledge-based approaches, the accuracy is largely affected by the quality of knowledge discovered through the data mining / machine learning system and implemented through an expert system.

Earlier knowledge-based geographical applications were based on decision trees and done by Hansen *et al*. (1996), Huang and Jenson (1997), and Friedl et al., (1999). Their studies demonstrated the great potential of using decision tree in spatial knowledge discovery. After them, the decision tree has been used intensively by many geographers including DeFries *et al*. (1998), Friedl *et al*. (1999), DeFries and Chan (2000), Pal and Mather (2001), Simard *et al*. (2002), Joy *et al*. (2003), McCauley and Goetz (2004), and Song *et al*. (2005). Their research suggested that the decision tree performs better than traditional classification technique. However, the problem of over-fitting seriously hinders its applications. When over-fitting occurs, the model cannot capture most testing data. One of the major factors causing an over-fitting is noisy data. Even without noisy data, the problem of insufficient training data can still cause an over-fitting problem. Although the decision tree has the capability to improve learning by itself, its learning is based only on the training data. A decision tree cannot enlarge the search space that may contain the model / knowledge capturing most testing data (Mitchell, 1997).

2.2. Land-Cover and Land-Use Change Detection

2.2.1. Concept of Land-Cover and Land-Use Change and Change Detection

LCLU change is a term linking both natural processes and human activities (Read and Lam, 2002). LCLU changes are usually measured as the differences in the states of LCLU features observed in different times or periods (Singh, 1989). There are the differences between LCLU conversion and LCLU modification. Four types of LCLU changes have been proposed: 1) feature change, 2) shape change, 3) location change, and 4) pattern change (Coppin *et al*., 2004). Feature change refers to LCLU conversion that is the complete replacement of one feature by another. The change of shape, location, and pattern refers to LCLU modification that affect the characteristics of the LCLU feature without changing its overall classification (Coppin *et al*., 2004; Khorram, 1999). Usually, LCLU modifications are more prevalent than LCLU conversions. LCLU usually changes continuously and catastrophically. Some changes are causes by human activities, such as urban growth and deforestation. Others change are caused by natural processes, such as wildfires and earthquakes. Many LCLU changes are triggered by both human activities and natural processes, such as wetland loss and range degradation (Coppin *et al*., 2004; Dolman and Verhagen, 2003; Lu *et al*., 2004b). These changes have substantially contributed to the increase in atmospheric CO_2, the exacerbation of water shortages, the changes of biogeochemical cycles, and the decrease of biodiversity at both regional and global scales (Dolman and Verhagen, 2003). During the last century, significant LCLU changes have heavily impacted human society. People are not only agents but also victims of LCLU changes.

2.2.2. Approaches to Land-Cover and Land-Use Change Detection

Change detection is one of the important components of LCLU study. It is a procedure to identify the difference in LCLU features observed within a given time period (Singh, 1989). In order to understand the relationships between human and nature better, and manage and utilize natural resources better, spatiotemporally and accurate LCLU change detection is extremely important and necessary (Lu *et al*., 2004a). Although the procedure to identify LCLU features at a particular time point and the procedure to identify LCLU features differences between or among different time points are clearly different, both LCLU classification and change detection are practically interwoven (Singh, 1989; Mas, 1999; Coppin *et al*., 2004; *Lu et al*., 2004a). A LCLU change detection includes at least three steps: 1) preprocess image (*e.g*., geometrical rectification, image registration, radiometric and atmospheric correction, and topographic correction), 2) apply suitable techniques to detect changes, and 3) assess accuracy. Logically, change detection is a procedure after data pre-processing, transformation, enhancement, segmentation, and classification.

The ability of any GIS system to detect LCLU change is a function of the "from-to" classes, and it deals with the spatial aspect, temporal aspect, and context of the change (Khorram, 1999). Not all interesting changes are easily detectable, and not all detectable changes are equally important (Coppin *et al*., 2004). A good understanding of the nature and the principles of LCLU change will enable the detection of the process rather than the detection of the change itself (Coppin *et al*., 2004). Although LCLU changes have been defined and measured as the differences between two or more end points within a given time period, more importantly, they are dynamic processes (Miller *et al*., 1998; Nagendra *et al*., 2004; Phillips *et al*., 2004; Petit *et al*., 2001; Turner, 1989). A systematic LCLU change detection should at least answer the following questions (Miller *et al*., 1998; Petit *et al*., 2001; Lu *et al*., 2004a):

- What type of LCLU features change and at what rate?
- Where and when do LCLU features change?
- How and why do LCLU change?
- What is the accuracy of LCLU change detection? and
- How do different factors (e.g., spectral, spatial, and temporal resolution, data sources, and data types, data layer combinations, and change detection techniques) impact it?

Since the 1980s, a number of change detection techniques have been developed (Table 2.2), and due to the importance of LCLU change detection, many new techniques have been

Table 2.2 A taxonomy of land-cover and land-use change detection techniques (Source: modified from Lu *et al*., 2004a; Coppin *et al*., 2004)

Categories	Characteristics	Examples	References
Algebra	Using math operators (e.g., subtract, divide, regression) on image data. It is easy to use but hard to interpret.	Image differencing	Sohl, 1999
		Image regression	Singh, 1986
		Image ratioing	Prakash and Gupta, 1998
		Vegetation index differencing	Lyon *et al*., 1998.
		Change vector analysis	Johnson & Kasischke, 1998
		Background subtraction	Singh, 1989
Transformation	Reducing data redundancy, but no detail information about LCLU change.	Principal component analysis	Parra *et al*., 1996
		Tasselled cap	Seto *et al*., 2002
		Gramm-schmidt	Collins & Woodcock, 1996
		Chi-square	Ridd and Liu, 1998
Classification	Providing detail information about LCLU change, but requiring large training data.	Post-classification comparison	Mas, 1999
		Spectral-temporal analysis	Soares & Hoffer, 1994
		Em detection	Serpico & Bruzzone, 1999
		Unsupervised change detection	Hame *et al*., 1998
		Hybrid change detection	Petit *et al*., 2001
GIS	Integrating methods & Incorporating data.	GIS/remote sensing integration	Price *et al*., 1992
		GIS approach	Taylor *et al*., 2000
Visual	Using experience, but time consumed.	On-screen digitizing	Sunar, 1998
		Visually interpretation	Slater and Brown, 2000
Artificial intelligence	Incorporating data, and discovering models, but hard to implement.	Artificial neural network	Liu and Lathrop, 2002
		Data mining	Mennis and Liu, 2005
		Knowledge discovery	Halid, 1997
		Expert system	Chalmers Fabricius, 2007
		Decision tree	Rogan *et al*., 2003
Processes	Generating process models, but not efficient	Multi-temporal analysis	Miller *et al*., 1998
		Logistic regression	Braimoh and Vlek, 2005
		Markov chain	Petit *et al*., 2001
Others	They are efficient for particular case, but not popularly used.	Spatial statistics approach	Read and Lam, 2002
		Change curves	Lawrence and Ripple, 1999
		Area production method	Hussin *et al*., 1994
		Indicators combination	Lambin and Strahler, 1994
		Structure-based approach	Zhang *et al*., 2002
		Li-Strahler reflectance model	Macomber & Woodcock, 1994
		Spectral mixture model	Adams *et al*., 1995
		Biophysical parameter method	Lu, 2001

proposed. Based on the studies done by Chan *et al*. (2001), Civco *et al*.(2002), Coppin *et al.* (2004), Lu *et al*. (2004a), Mas (1999), and Singh (1989), most existing LCLU change detection techniques could be categorized into the following groups:

- Algebra-based approaches use algebra as a key algorithm for change detection. Algorithms, such as image differencing, image regression, image ratioing, change vector analysis, vegetation differencing, and background subtraction, are most commonly used change detection techniques. Algebra-based approaches, except change vector analysis, are easy to implement. Although their results are easy to interpret, these approaches do not provide detail information about LCLU change.
- Transformation-based approaches include principal components analysis (PCA), tasseled cap transformation (TCT), Gram-Schmidt, and Chi-square transformations. These approaches are usually to perform data preprocessing. After transformation, the resultant data will be used for LCLU change detection. The common advantage of these methods is data redundancy reduction. However, the resultant information is not explicit. These approaches require the selection of thresholds, and the results are difficult to interpret.
- Classification-based approaches include post-classification comparison, spectral-temporal combined analysis, expectation-maximization algorithm, unsupervised change detection, and hybrid change detection. Although these approaches are able to provide some detailed information about LCLU change, they require high-quality and sufficient-quantity of training samples.
- GIS-based approaches integrate GIS and remote sensing methods for the purpose of change detection, so that data from different sources or with different formats can be incorporated into the procedure of change detection. These approaches may improve

or hinder the performance of change detection based on data types and algorithms used (Huang and Jenson, 1997). The key idea is to introduce more information into the procedure of LCLU classification and change detection.

- Visual analysis-based approach includes visual interpretation of multi-temporal image composite and on-screen digitizing of change areas. The advantage is that an analyst's experience and knowledge can be fully used, but it consumes a lot of time and is hard to update. The quality of result will be heavily impacted by the quality of analyst's knowledge.
- Artificial intelligence (or machine learning, or data mining, or knowledge discovery) based approaches include artificial neural network, fuzzy logic, decision tree, learning vector quantization, association rule, genetic algorithms, and so on (Chan *et al*., 2001; Liu and Lathrop, 2002). These methods are not based on the statistical assumption. They can work with different data sources, and usually perform better than traditional methods. The drawback is that they need high quality and quantity of training data.
- Process-based approaches include multi-temporal analysis, multiple linear regression, and Markov chain. These methods can provide some information about LCLU change processes, so that the causes of change could be inferred. However, the existing multi-temporal analysis approach is not efficient (Yang and Lo, 2002), both multiple regression and Markov chain-based approaches assume that the LCLU change processes are stochastic (Petit *et al*., 2001). Unfortunately, this is not always true.

Although a large number of change detection techniques have been proposed and tested so far, it is improper to argue which approach is best suitable for a particular study area. There is no single suitable method for all cases. The method selected depends on the user's experience of handling remote sensing data and ancillary data, their knowledge of LCLU change detection methods and the study area, and their capability of accessing different resources (Lu *et al*., 2004a). Integration of different approaches and incorporation of various data may improve the performance of change detection. Previous research has shown that most change detection research can partially or fully answer the questions, such as which land features change and where, when, and at what rate. However, most of them could not provide information on how and why LCLU changes. To answer these questions, the process-oriented change detection is necessary.

2.2.3. Processes-Oriented Land-Cover and Land-Use Change Detection

LCLU changes are dynamic processes. Process-oriented LCLU change detection not only provides information about the features, patterns, and rates but also provides information about the processes. It gives more detailed information about the changes. Although models based on stochastic assumption have been applied successfully in LCLU change process studies (Guerra *et al*., 1998; Lambin, 1996; Mertens and Lambin, 2000; Petit *et al*., 2001; Singh, 1989; Turner, 1987 and 1989), those models may not be the best approaches for LCLU change processes modeling. An alternative approach is to discover the knowledge of change process through spectral-spatial-temporal data mining, and then use the resultant knowledge to model the LCLU changes.

Process mining or workflow discovery is an approach to discover process knowledge. The last decade has experienced a tremendous development in process mining and workflow discovering in business and manufactory management area (Dustdar *et al*., 2005; Reddy *et al*., 2001). Its major idea is to automatically extract knowledge and build a model that describes the behavior of system from event logs recorded by an information system or compiled by users (van der Aalst and Weijters, 2005; Alves De Medeiros *et al*., 2007). Among the major language-dependent process mining / workflow discovery techniques (*e.g.*, unified modeling language, event-driving process chains, and Petri net), Petri net, a popular, powerful, and easily implemented technique, has been developed from the early work of Carl Adam Petri (Peterson, 1981; Wu, 2006). Petri net, namely place / transition (P/T) net, is both a graphical language and a mathematical structure, and it has been widely used to build and simulate the discrete event systems (Desel, 2005; Wu, 2006). A popular Petri net-based process mining application is implemented through the α-Algorithm that transforms an event log file into a Petri net. Unfortunately, this algorithm has at least three limitations: 1) invisible tasks, 2) non-free-choice, and 3) duplicate tasks (van der Aalst and Weijters, 2005). Since the knowledge of processes is derived only based on the provided training data, the quality of training data will impact the quality of knowledge. Moreover, two drawbacks have seldom been pointed out by previous research: limited search space and non-global optimization. GAs provide random search and global optimization mechanism that could overcome these limitations existing in a pure Petri net approach.

2.3. Land-Cover and Land-Use Predictive Modeling

2.3.1. Concept of Land-Cover and Land-Use Predictive Modeling

LCLU changes are driven by human activities, natural processes, or the combination of both. They can be abstracted or simplified by various models at different spatial and temporal scales, and with different accuracies. LCLU models can be defined as an abstraction or simplification of real-world LCLU dynamics. They can be used to explore the dynamics and drivers of LCLU change and inform policies affecting such change. The modeling procedure of LCLU includes 1) collecting and processing data with the consideration of spatial, temporal, spectral, and radiometric resolution, 2) selecting or constructing the suitable approaches with the consideration of uncertainty and feasibility, 3) calibrating and validating the models with the consideration of complexity and accuracy, and 4) implementing and evaluating the optimized models with the consideration of LCLU prediction and reconstruction (Wu, 2002; Candau *et al.*, 2000; Candau, 2002;). Since predictive modeling provides a foundation for regional or local sustainable development planning, and reconstruction modeling only provides information about the past LCLU, this research only focuses on LCLU predictive modeling.

2.3.2. Approaches to Land-Cover and Land-Use Predictive Modeling

LCLU predictive models not only offer the possibility to test the sensitivity of LCLU patterns to the changes of selected variables, but also allow testing of the stability of linked human and natural systems (Veldkamp and Lambin, 2001). Since the 1980s, a number of LCLU modeling techniques have been developed (Agarwal *et al*., 2002; Jones, 2005; Lambin, 2004; Loveland *et al*., 2000; Parker *et al*., 2001 and 2003; Singh, 2003; U.S. EPA, 2000; Waddell, 2002; Zhao and Chung, 2006). Based on the classification schemes used by the above scholars, those LCLU modeling techniques can be categorized into eight groups (Table 2.3): 1) empirical / statistical model, 2) stochastic / discrete model, 3) spatial interaction / input-output model, 4) linear programming / multinomial logic model, 5) rule-based model, 6) process-based model, 7) cellular automata / agent-based model, and 8) hybrid / other model. Some models are developed based on different techniques. If a model employs various algorithms or approaches, it will be classified based on the approach that the model mostly emphasizes.

- The empirical-statistic models are commonly used approaches in LCLU modeling. They are based on the multiple linear regression techniques and geospatial statistical methods (Ludeke *et al*., 1990). This type of models can be used to identify explicitly the associated relationships among different LCLU features using multi-variants analysis of possible exogenous contributions to empirical derived rates of changes (Lambin, 2004). However, this approach can be used only to model statistical association, and any causal relationship will not be found.

Table 2.3 A taxonomy of land-cover and land-use change predictive modeling techniques (Source: modified from Agarwal, 2002; U.S. EPA, 2000; and Zhao *et al*., 2006)

Categories	Characteristics	Examples	References
Empirical and Statistic-based	Using multiple linear regression, and explain changes in a statistical sense.	CLUE	Veldkamp and Fresco, 1996a and 1996b
		CUBRA	Landis & Zhang, 1998a and 1998b
		CUF, CUF II	Landis, 1995
		LTM	Pijanowski *et al*., 2002
Stochastic model-based	Markov chains, stationary transition, short term prediction.	LUCAS	Berry *et al*., 1996
		Markov Chain	Muller & Middleton, 1994
		METROSIM	Anas, 1994
Spatial interaction and input-output-based	Using gravity theory, population distribution is the function of places; attractiveness; also using input-output model in economics, too address spatial pattern of society.	DRAM / EMPAL	Putman, 1995
		LILT	Mackett, 1990a and 1990b
		HLFM II+	Dowling *et al*., 2000.
		LUTRIM	Mann, 1995
		MEPLAN	Hunt, 1997
		TRANUS	U.S. EPA, 2000
		DELTA	U.S. EPA, 2000
Lnear programming and Multinominal logit based	Using linear, nonlinear, dynamic, hierarchical, and goal programming.	HerbertStevensModel	Herbert & Stevens, 1960
		TOPAZ / TOPMET	Zhao and Chung, 2006
		POLIS	Zhao and Chung., 2006
Rule-based	Based on economic theories and market rules, use for long-term scenario at county level.	CUF-1 / CUF-2	Landis, 1994
		SAM / SAM-IM	U.S. EPA, 2000
		UPLAN	Johnston *et al.,* 2003
		What If	Klosterman, 1999
		SLAM	Zhao and Chung , 2006
		ULAM	Zhao and Chung , 2006
Process-based	Process-based, good for representing non-stationary processes.	PLM	Voinov *et al*., 1999b
		GEM	Fitz *et al*., 1996
		IMAGE	Alcamo *et al*., 1998
		IMPEL	Rounsevell *et al*., 1997
Cellular Automata and Agent-based	Cell/agent, flexible, bottom up, good for LCLU prediction, hard to define rules or behaviors.	SLEUTH	Herold *et al*., 2003
		CA/LUCC	Messina, 2001
		ABM/LUCC	Parker *et al*., 2001
		MAS/LUCC	Parker *et al*., 2003
Hybrid and Other	Integrate one model with another model or GIS. It uses both strength and overcomes drawback.	DT-CA	Li & Yeh, 2004
		ANN-CA	Li & Yeh, 2002
		SmartGrowth INDEX	U.S. EPA, 2007
		SmartPlaces	EPRI, 2002
	Use model to simulate the individual behaviors, apply conclusion to entire population.	UrbanSim	Waddell & Ulfarson, 2004
		IRPUD	Wegener, 1998
		MASTER	Mackett, 1990a, 1990b
		NBER/HUDS	Kain, 1986

- The stochastic models are based mainly on transition probability models such as Markov chains. In this type of approaches, the LCLU change processes are stochastically described by a set of states and steps. The states of the system are defined as the pattern of LCLU features, and the transition probabilities are estimated statistically from a sample of transitions occurring during time intervals (Lambin, 2004). Such models rely on the assumption of a stationary transition matrix. Unfortunately, this assumption is not valid in most cases. This type of models can be used only for predicting short-term LCLU change under a strict assumption of stationary of process.
- Spatial interaction models and spatial input-output models are based on the gravity theory and the theory of input-output respectively. Spatial interaction models are designed to link the pattern of population distribution, the attractiveness of places, and the travel cost (Putman, 1995; Mackett, 1990a and 1990b; Dowling *et al*., 2000). Such models can represent only limited spatial detail (Waddell and Ulfarsson, 2004). Spatial input-output models are developed to address spatial patterns of economic activities within the various regions and the spatial pattern of the flow of money, goods, and the movement of people among the different places (Hunt, 1997; U.S. EPA, 2000). Such models are expressed by static equilibrium solution (Waddell and Ulfarsson, 2004). However, in most cases the processes and relationships among the different components of LCLU system are dynamitic.
- Optimization models are based on linear programming, multinomial logic, or general equilibrium. They are developed and used in economics-related areas (Lambin, 2004). In this type of models, LCLU change is a function of choices by landowner among various rents because a land parcel with the given attributes and location is

treated as being used in the way that generates the highest rent. In other words, the behavior of landowner is described with the models. The optimized models assume that the landowner can make an informed LCLU prediction. However, the applicability of such models for LCLU changes prediction is limited because of unpredictable economic factors (*e.g.*, price fluctuations and demand market) and non-economic factors (*e.g.*, natural hazard). They also suffer from other limitations, such as the arbitrary definition of objective functions and non-optimal behavior of people (Lambin, 2004).

- The process-based models are developed to simulate and analyze the interactions among different system components (Lambin, 2004). LCLU changes are a set of processes. Process models can be used to predict temporal changes in spatial patterns of LCLU when they are spatially and temporally explicit. The strength of a LCLU process model depends on whether the major features affecting LCLU change are integrated, whether the functional relationships among factors affecting change processes are represented appropriately, and whether the model is able to predict the most important LCLU changes (Lambin, 2004). Process models are well-suited to representing non-stationary process because they mimic the underlying processes in the system.
- CA-based models and agent-based models are similar. Both of them are "bottom to top" approaches. A cellular automaton has at least five components: space, state, time step, neighborhood, and transition rules. Its basic unit is a cell. Cells change from one state to another state based on the configuration of those five components (Gaylord and Nishidate, 1996). An agent-based model consists of autonomous decision-making entities (agents), an environment through which agents interact, rules that define the

relationship among agents, and rules that determine sequencing of actions in the model (Parker *et al*., 2003). In LCLU modeling, the behavior of cell or agent will determine the resultant LCLU patterns. These approaches are very flexible. The major drawback is the difficult of developing the cells' transition rules or defining an agent's behaviors.

- Hybrid models combine at least two different techniques together, such as the integration of a decision tree and CA (Li and Yeh, 2004). The purpose of hybrid models is to implement the strength from different approaches or to use one approach's strength to overcome another approach's drawback. In CA-based LCLU modeling, defining transition rules was a difficult work, but it can be performed easily by using a decision tree. Hybrid models provide more approaches for LCLU predictive modeling.

Although a large number of predictive modeling techniques have been developed and tested, it is difficult to state which approach is most suitable for a specific study area. Currently, the most commonly used predictive modeling is CA-based approaches, and the major reason can be traced to its simplicity and modeling capability. However, developing a proper transition rule set remains a challenge (Li and Yeh, 2004).

2.4. Genetic Algorithms

2.4.1. Concept of Genetic Algorithms

Genetic algorithms (GAs) were proposed as a type of efficient stochastic search and global optimization technique by both Holland (1975) and Goldberg (1989). GA-based approaches seem to have great promise because of its being inspired by the principle of genetics and Darwin's evolutionary theory. Unlike other algorithms (such as neural network and fuzzy set), GAs can automatically enlarge search space and optimize results. The algorithms begin with a randomly initialized population and evolve toward better solutions by using fitness evaluation and genetic operators. In the process of evolution, individuals with higher fitness value have greater opportunities to reproduce offspring. The fitness-based natural selection is balanced by randomly adding genetic operations through crossover and mutation (Wu and Cao, 1997). The whole course is controlled by a number of genetic parameters discussed below.

2.4.2. Applications of Genetic Algorithms

In GAs, the representation of individual is an important issue. For example, when using GAs to solve a problem, people are usually interested in a set of solutions, rather than a single solution. There are two approaches to encode a set of solutions in a GA population. In the conventional GA approach, each individual of GA population represents a set of solutions, or entire candidate solutions. This is called the Pittsburgh approach. Another approach, which departs from conventional GA, consists of having an individual represent a single solution, or a part of a candidate solution. This is called the Michigan approach (Freitas, 2002). In this dissertation, the Pittsburgh approach-based GA software was used because of availability.

Since the late 1980s, genetic algorithms have been applied successfully to various geographical studies, such as 1) image processing, feature extraction, and classification; 2) integrating or comparing with different approaches for geospatial research purposes; 3) spatial optimization and regional planning; and 4) mapping and others (Table 2.4).

2.4.3. Genetic Parameters

GAs have five types of parameters including: number of generations, population size, crossover rate, mutation rate, and generation gap or elite rate. Generation is a procedure to generate a new population, population is a group of individuals with same or similar characteristics, crossover is a procedure to generate children from two parents, mutation is a procedure to change selected individual genome, and generation gap is the replacement rate of

Table 2.4 Summary of genetic algorithms-based geographical applications (Source: modified from Krzanowski and Raper, 2001)

Categories	Characteristics	Applications	References
Image processing, features extraction, and classification	Genetic algorithms are used for processing image, extracting features, and classifying image data.	Feature extraction	Brumby *et al*., 1999
		Image processing	Harvey *et al*., 2000; Brumby *et al*., 1999
		Classify multi-sources data	Tso & Mather, 1999
		Features classification	Perkins *et al*., 2000
		Pixels classification	Bandyopadhyay & Pal, 2001
		Image classification	Stathakis &Vasilakos, 2006
		Road detection	Jeon *et al*., 2002
		Sharpen multi-spectral image	Garzelli & Nencini, 2006
		Sub-pixel mapping	Mertens *et al*., 2003
Integration / comparison	Genetic algorithms are used to integrate or compare with neural network, fuzzy logic, or GIS.	GA-ANN / spectral identify	Clark & Canas, 1995
		GA-ANN / decision making	Zhou & Civco, 1996
		GA-ANN / feature extraction	Zhou & Civco, 1997
		GA-Fuzzy / classification	Maulik & Bandyopadhyay, 2003
		GA-Fuzzy / classification	Bandyopadhyay, 2005
		GA / standard methods	Pal *et al*., 2001
		GA/ PCA	Garcia-Orellana *et al*., 2002
		GA / supervised classifiers	Harvey *et al*., 2002
		GA / ANN / Fuzzy	Demetris & Vasilakos, 2006; Kulkarni and Lulla, 1999
		Integrating with GIS	Stockman *et al*., 2006
Spatial optimization, and planning	Genetic algorithms are used to optimize various geospatial solutions	Location allocation	Hosage & Goodchild, 1986
		Location optimization	Li & Yeh, 2005
		Site-search problems	Xiao *et al*., 2002
		Route planning	Huang *et al*., 2004
		Route and road design	Pereira, 2001
		Land use planning	Stewart *et al*., 2004
		Land use planning	Matthews *et al*., 2000
		Airspace sectoring	Delahaye, 2001
		Policy's geo-consequence	Bennett *et al*., 2004
		Solve various GIS problems	Dijk *et al*., 2001
		Optimize patch configuration	Brookes, 2001
		Photogrammetric network	Olague, 2002
		Optimize sensor model	Samaadzadegan *et al*., 2005

Table 2.4 (Continued)

Categories	Characteristics	Applications	References
Spatial modeling, analysis, and prediction	Genetic algorithms are used for modeling, analyzing, and predicting.	Spatial modeling	Openshaw, 1988, 1992,1995
		Spatial modeling	Krzanowski & Raper, 2001
		Spatial modeling	Dibble & Densham, 1993
		Spatial reference analysis	Cooley *et al.*, 1997
		Spatial analysis	Dibble, 2001
		Species distribution model	Stockwell & Peters, 1999
		Spider distribution analysis	Bond *et al.*, 2006
		Species distribution model	Stockman *et al.*, 2006
		Species biodiversity model	Stockwell *et al.*, 2006
		Mammals distribution model	Patricia *et al.*, 2004
Map Generalization	Optimizing map display.	Map generalization	Dijk *et al.*, 2002
		Class intervals on maps	Armstrong *et al.*, 2003
Others	Improve environmental monitoring.	Land roughness information	Jin & Wang, 2001
		Evaluate water quality	Chen, 2003
		Assess ground water quality	Armstrong & Bennett, 1990

each generation. Although the impact of genetic parameters on the performance of genetic algorithms has been studied in other areas (Grefenstette, 1986), it is seldom investigated in geospatial analysis area.

In spite of great achievement, GA-based approaches still suffer from a number of problems. Although being proposed as a random search and optimization technique, GAs never guarantee global optimum because both inheritance and evolution are very complex processes and controlled by many different factors. The most harmful one is that any improper setting of genetic / image parameters could cause the population lose its diversity, be dominated by so-called elites, or prematurely converge to local optimum (Wu and Cao, 1997; Rocha and Neves, 1999). In this type of situation, the search and optimization procedure is trapped in the sub-optimal state and most of the operators cannot produce offspring surpassing their parents (Liu *et al.,* 2000). It has been proven that genetic parameters have much influence on the performance of genetic algorithms (Yang *et al*., 2000), and classical genetic algorithms cannot converge to the global optimal solution without properly configuration of genetic parameters (Rudolph, 1994).

The last decade has experienced extensive applications of genetic algorithms in geographical research. Based on the various research, a literature review is developed (Table 2.4), and it suggests that:

- Existing studies have been dominated by applications of GAs in satellite image analysis, location / route optimization, mapping optimization, comparison/integration of different approaches, and spatial modeling. Most studies have been done since the mid-1990s.
- Systematic and scientific investigations on premature convergence / local optimum problem in GAs-based geographical studies, and the relationships among the performance of GA, genetic / image factors, and the accuracy of LCLU have been

scarce. It is a challenge to find a general and effective approach to improve the performance of GAs in geographical studies.

2.5. Petri Nets

2.5.1. Concept of Petri Nets

A LCLU system can be regarded as a discrete event system, the events are LCLU changes, and the states are a set of feature classes. State-to-state transitions are driven by human activities, natural process, or a combination of both. The behavior of discrete event systems can be captured and analyzed by using modeling techniques. Petri nets are the ideal techniques for such modeling because they provide both graphical representation and mathematical representation. The concept of the Petri nets was introduced by Petri, C. A. in his dissertation in 1962 (Schneeweiss, 2004). Petri nets are a type of stream models (Liu and Hao, 2005), they are also a type of graphical representations and mathematical modeling techniques for describing analyzing, and modeling the systems' temporal procedures (or dynamic processes) (Schneeweiss, 2004), and the systems are characterized as being concurrent, asynchronous, distributed, parallel, discrete, nondeterministic, and / or stochastic (Murata, 1989). As a graphical technique, Petri nets can be used as a visual-communication aid similar to flow charts, block diagrams, and / or networks. As a mathematical approach, Petri net provides the possible to set up equations or mathematical models to manage, analyze, model a system's behavior (Murata, 1989).

2.5.2. Applications of Petri Nets

A classical Petri net is composed of places, tokens held by places, transitions, and directed arcs between the places and transitions (David and Alla, 1992; Schneeweiss, 2004; Wu, 2006). The number of tokens is called the marking, and the distribution of tokens among the places represents the state of the Petri net. The arcs determine how tokens move from one place to another place upon the firing of a transition. The state of Petri net can be represented as an integer vector, and the architecture or layout of a Petri net can be represented with an integer matrix known as the incidence matrix (David and Alla, 1992; Schneeweiss, 2004; Wu, 2006). Based on a classical Petri net theory, at least three important types of Petri net can be developed. First, timed Petri nets include time in the net by associating a time delay with the firing of a transition. Second, colored Petri net has colored tokens. Last, generalized Petri net has weights associated with the arcs (David and Alla, 1992; Schneeweiss, 2004; Wu, 2006). Those types of Petri nets are commonly used in PN-based application. They are closely related to this dissertation study.

Due to their generality and permissiveness, Petri nets have been proposed for various applications. Graphically, they can be used for any system having discrete, parallel, or concurrent behavior. However, mathematically, serious attention must be paid to the balance of modeling generality and analysis capability. The major users of Petri nets are computer and automatic control scientists (David and Alla, 1992). Petri nets have been used in other areas. Based on the purposes, the applications can be categorized into four groups: 1) industrial processes control, 2) business processes management, 3) computer networking or communication, and 4) scientific research (Table 2.5).

2.5.3. Limitation of Pure Petri Nets

A literature review indicates that manufacture processes control (Department of Computer Science, University of Aarhus, 2007), business processes management (Desel *et al.,* 1998), and computer science study (Murata, 1989) are major application fields. Although Petri nets have been introduced into many other scientific research, such as in biology and chemistry (Will and Heiner, 2002), there are very few Petri net applications in geographic research. Traditionally, the process models generated by Petri net are based on the input data only. If the space of the input data is not large enough, the resultant process model cannot fit the entire data

Table 2.5 Summary of Petri nets applications

Categories	Characteristics	Applications	References
Manufacture / industry processes control	Use Petri nets to extract and model industrial process	Discrete event system	Holloway *et al*., 1997
		Manage manufacture system	Lambin & Strahler,, 1994
		Industrial system control	Suri, 1985
		Traffic control	Hsieh and Chen, 1999
		Industrial system modeling	Rodrigo and Nicholls, 1998
		Industrial processes	van der Aalst, 1994
		Manufacture management	Wu *et al.,* 2006
		Re-configurable system	Kumar et al., 2005
		Production schedule	Chien and Chen, 2007
Business processes management	Use Petri nets to extract and model business process	Manage business processes	van der Aalst *et al*., 1996
		Virtual enterprises	Gou *et al*., 2000
		business processes	van der Aalst, 2002, 2003
		Web service composition	Hamadi & Benatallah, 2003
		Verification of Workflow	van der Aalst, 1997, 1998
		Process mining	van der Aalst *et al.* 2004, 2005
Computer science and technology	Use Petri nets to extract and model computer network structure	Multi-processor system model	Gourgand, 1993
		Communication protocol	Jaragh and Saleh, 1999
		Software development	Chang *et al*., 1989
		Distributed system modeling	Yi and Kochut, 2005
		Parallel programming	Ferscha, 1992; Yen, 2002
		Concurrent processes	Bulitko and Wilkins, 2003
Scientific research	Use Petri nets to extract and model natural process	Spatio-temporal modeling	Hsu *et al*., 2003
		Spatio-temporal modeling	Wang & Nakayama, 2005
		Biological research	Doi *et al*.,1999; Hirata, 2001
		Ecological research	Gronewold & Sonnenschein, 1998
		Chemistry research	Kuroda and Ogawa, 1994
		Medicine research	Peimann, 1988
		GIScience research	Yin & Li, 2004
		Geospatial Modeling	Liu and Hao, 2005

set. Optimization may be an approach to solve this problem. One common limitation in pure Petri nets applications is the lack of efficient optimization.

2.6. Cellular Automata

2.6.1. Concept of Cellular Automata

Faced with the problem of constructing a mathematical model of LCLU change pattern, the first modeling step is to clarify the level of CA structure in which one is primarily interested with regarding to space, time, state, and interactions (Deutsch and Dormann, 2005). One way to classify the approaches of spatio-temporal modeling is to distinguish the difference between continuous and discrete state, time, and space variables. A classification of different approaches is shown in Table 2.6.

CA is a type of dynamic system that is discrete in terms of time, state, and space. CA plays an important role in spatio-temporal modeling. Originally, CA were developed by John von Neumann and Stanislaw Ulam in the 1940s for investigating the behavior of complex systems (White and Engelen, 1993). Both of them attempted to develop a theory of machine behaviors that can reproduce the structure by itself. The concept of a self-organizing system is the key of CA (Mortlock, 2004). A cellular automaton includes five components: space, state, time step, neighborhood, and transition rule. Space is composed of individual cells, and each cell may change from one state to another state based on the transition rules, their neighborhood, and time steps (Gaylord and Nishidate, 1996). The structure of CA tends to develop naturally ordered patterns on a large scale (Torrens, 2000; 2001). Since being developed, CA had not become popular until 1970, when *The Game of Life*, which combined all the components of CA in a model and simulated the key elements of reproduction in simple way, was invented by mathematician John Conway and published by Martin Gardner. However, much influence on later development of CA theories, techniques, and applications can be attributed by Wolfram's

Table 2.6 Characteristics of cellular automata modeling approaches
(Source: Deutsch and Dormann, 2005)

Model Approaches	Space Variable	Time Variable	State Variable
PDEs, integro-differential eqs	Continuous	Continuous	Continuous
Spatial point process set or rules	Continuous	Continuous	Discrete
Integro-different eqn.	Continuous	Discrete	Continuous
Set of rules	Continuous	Discrete	Discrete
Coupled ODEs	Discrete	Continuous	Continuous
Interacting particle systems	Discrete	Continuous	Discrete
Coupled map lattices, system of diff. eqns, lattice-Boltzmann models	Discrete	Discrete	Continuous
Cellular automata, lattice-gas cellular automata	Discrete	Discrete	Discrete

research in 1984. He demonstrated that the origins of the complexity of the natural system could be investigated through CA. Since then, the CA-based geographic research have dramatically increased.

2.6.2. Applications of Cellular Automata

The introduction of CA in geographic research can be traced back to the work done by Hagerstrand (1967; 1968). In his diffusion models, he highlighted the major components of the current CA structure: cell, lattice, state, neighborhood, time, and transition rules. Although his research was theoretically well-defined, it was practically limited by the simulation capability. Tobler (1979) may be the first one to clearly suggest the use of CA in geographic research. He developed a set of CA-based spatial models to study land use. The work done by Tobler laid the theoretical foundation for the current studies on geographic CA. However, the temporal dimension of his CA model was not strong (Wegener, 2000). Since the mid-1980s, the works done by both Hagerstrand and Tobler have been improved theoretically, methodologically, and practically by CA-based geographic research. The research done by Couclelis (1985, 1987, 1988) may be most significant. She adopted CA as a simulation framework in which the local and global behaviors interact. She also explored a classical 2-D CA model with a detailed description of each of the CA components. Urban growth has been attracting most attention from the geospatial modeling community. A series of papers on urban growth, published by Batty (1991; 1998; 2000; 2002), Batty and Xie (1994; 1997), Batty *et al.* (1989; 1999), Clarke and Gaydos (1998), Clarke et al (1994, 1997), Landis (1994, 1995), Landis and Zhang (1998a, 1998b), and White and Engelen (1993, 1994, 1997, 2000), have demonstrated very important insight into the nature and form of urban and regional land use dynamics. CA have also been used in other geographic research. Based on the issues studied, the CA-based geographic research can be categorized into the following groups: 1) urban growth, 2) regional LCLU change, 3) population dynamics, 4) disease spread, 5) wildfire propagation, and 6) biological distribution (Table 2.7). One aspect should be mentioned is agent-based approaches in geographic research. Essentially, agent-based approaches are developed from CA, both of them are pretty much same.

2.6.3. Transition Rules in Cellular Automata

Existing studies demonstrate that the most important component in geographic CA is the transition rule. The transition rule is the key component of CA. The real-world processes or behaviors are translated into a geographic CA model through transition rules only. Transition rules drive and guide CA dynamic evolution. The definitions of transition rules in geographic CA are strongly based on domain knowledge and individual experiences. For example, the transition rules in CA-based urban modeling are usually given according to the intuitive understanding of the process of urban growth. Transition rules can be defined using various mathematical expressions, such as nested neighborhood spaces and distance decay functions (Batty and Xie, 1994), predefined parameter matrices (White and Engelen, 1993), linear equations of multi-criteria evaluation (MCE) (Wu and Webster, 1998), logistic models (Wu, 2002), grey-cell or fuzzy states (Li and Yeh, 2000), and neural networks (Li and Yeh, 2002). However, most transition rules are not easy to understand because they use mathematical equations instead of explicit human understandable language. Li and Yeh (2004) used the decision tree to discover a set of explicit transition rules for a geographical CA. But decision tree approach has over-fitting problem, and it has no capability of global optimization. Although the number of ways of defining transition rules seems to be virtually unlimited, how to discover and optimize transition rules in an objective way requires further study.

In addition to defining transition rules, the calibration of geographical CA seems to be very difficult also, because a large number of rules have been used. Usually, transition rules

Table 2.7 Summary of cellular automata applications

Categories	Characteristics	Applications	References
Urban	CA-based urban growth modeling	Urban modeling	Batty,1998, 2000
		Urban modeling	Clarke *et al*., 1997
		Urban modeling	Landis & Zhang, 1998a, 1998b
		Urban modeling	White & Engelen, 1994, 1997, 2000
		UrbanSim	Waddell, 2002
		Urban modeling	Xie, 1996
		Urban growth modeling	Chen *et al*., 2002
		Urban planning	Yeh & Li, 2000, 2001, 2003
		Space-Time processes	Wu, 1999
Land	CA-based land-cover and land-use change modeling	Land use change	White & Engelen, 1993
		Land use change	Batty & Xie, 1994
		Land conversion simulation	Wu, 1998a, 1998b
		Land-cover and land-use change	Singh, 2003
		Land-cover and land-use change	Parker *et al*., 2003
		Land use change	Li & Yeh, 2002,
		Land use change	Jenerette & Wu, 2004
Population	CA-based population migration modeling	Population density	Yeh and Li, 2002
		Population surface modeling	Wu & Martin, 2002
		Population spatial spread	Songderath & Schroder, 2004
Disease	Ca-based disease spread modeling	Epidemic dynamics	Willox *et al*., 2003
		Influenza A viral infections	Beauchemin *et al*., 2005
		Epidemic propagation	Sirakoulis *et al*., 2000
		HIV infection dynamics	Dos Santos & Coutinho, 2001
		HIV infection	Sloot *et al*., 2002
Wildfire	CA-based wildfire propagation and extinction modeling	Propagation and extinction	Clarke *et al*., 1994.
		Predict wild bush fire spread	Mrax *et al*., 1999
		Simulate savanna wild fire	Berjak & Hearne, 2002
		Predit forest fire spread	Karafyllids & Thanailakis, 1997
Biology and Ecology	CA-based vegetation, animal, and insect, and habitat distribution modeling	Species distribution	Carey, 1996
		Vegetation dynamics	Balzter *et al*., 1998
		Simulate forest pattern	Green *et al*., 1985
		Invasive species spread	BenDor & Metcalf, 2006
		Habitat pattern	Syphard *et al*., 2005
		Deforestation	Messina and Walsh, 2000
		Plants dispersal	Harada & Iwasa, 2006

consist of many variables and parameters. There are many uncertainties in determining parameter values. Urban CA is very sensitive to transition rules and their parameter values (Wu and Webster, 1998; Li and Yeh, 2002; Wu, 2002). There are very limited studies on the calibration of geographical CA (Clarke *et al*., 1997; Clarke and Gaydos, 1998; White and Engelen, 1997; Wu, 2002), and most of them are the so-called "trial and error" approaches.

Based on the above discussion, the CA-based urban growth modeling is very complex. When we attempt to develop a regional LCLU model, there are many land features involved simultaneously. The complexity of geographical CA will increase tremendously. The literature review indicates that CA-based LCLU predictive modeling still has at least following three problems: 1) how to develop a set of optimized objective transition rules, 2) how to calibrate and validate a geographical CA, and 3) how to identify the effects of CA's components on their performance.

2.7. Summary

This chapter discussed the major advances in LCLU classification, change detection, predictive modeling, genetic algorithms, Petri nets, and cellular automata, which can be summarized as follows:

- During the last two decades, the LCLU classification techniques have been shifted from spectral analysis to contextual, spatial syntactic analysis, and to knowledge-based interpretation. Several knowledge-based approaches, such as the decision tree, fuzzy logic, artificial neural network, genetic algorithm, have been used in the procedure of LCLU classification. Different approaches have different capabilities in terms of random search and global optimization.
- During the last two decades, the LCLU change detection techniques have been shifted from algebra-based approaches to artificial intelligence-based approaches, from two

or several endpoint differences detection to process-oriented detection. Since LCLU changes are dynamic processes, a systematic LCLU change detection should be process-oriented. However, currently, most LCLU change detections are not process-oriented.

- During the last two decades, the LCLU predictive modeling techniques have been shifted from statistic / stochastic approaches to dynamic / cellular approaches. Both process models and transition rules are crucial for the new approaches. However, developing a set of process models or transition rules for representing a LCLU system remains very challenging.
- Genetic algorithms have been used in geographic research since the late 1980s, and most applications are about LCLU classification and spatial modeling. However, their intrinsic premature convergence problem and its impacts on the accuracy of LCLU classification or spatial modeling have seldom been investigated.
- Petri nets have been used in the management of manufacturing process, business process, and computer network. They are relatively new in geographic research and application. Based on the existing studies in various scientific fields, most current applications use pure Petri net, and the resultant process models are not optimized. In the dissertation research, a genetic Petri net is involved, and the resultant process models are optimized.
- Cellular automata have been used in geospatial modeling for over two decades. The development of transition rules remains a problem. Recently, a data mining based approach for transition rules development was proposed, but those transition rules were not optimized. Also, the impacts of space, state, time step, neighborhood, and transition rule on the performance of geographic cellular automata need further study.

CHAPTER 3 RESEARCH METHODOLOGY

3.1. Study Area and Data Sources

3.1.1. Natural and Social Environment

Louisiana has twelve water management basins delineated by natural drainage patterns of the state's major river basins. The study area, namely Tickfaw River watershed in the upper Lake Pontchartrain Basin, is especially challenged from rapid population growth, industrial activities, and agricultural use. This watershed is located in southeastern Louisiana and southwestern Mississippi, between 30˚19'14" - 31˚9'42" N latitude and 90˚31'58" - 90˚50'15" W longitude (Figure 3.1). The Tickfaw River flows from the Mississippi state line to Springville at Louisiana Highway 42 then to the Lake Maurepas. Its drainage area totally covers 711 $mile^2$ and includes a portion of Amite County, St. Helena Parish, Livingston Parish, and Tangipahoa Parish (U.S. EPA, 2005).

Southeastern Louisiana is the wettest part of the state, and its climate is classified as humid subtropical. The Gulf of Mexico dominates the climate by providing a flow of warm and humid air into this area (Johnson and Yodis, 1998; Yodis *et al.,* 2003). Based on a long-term research done by Wu (2005), the annual average air temperature during 1948-2000 is about 19°C, with the lowest monthly average of 12°C in January and the highest monthly average of 28°C in July. Although monthly average precipitation is well-distributed throughout the year, autumn tends to be the driest season with a distinct minimum in October. Much of the winter rainfall is associated with the mid-latitude wave cyclones. During the summer months, local heating produces summer thunderstorms, which can occur on a near-daily basis (Johnson and Yodis, 1998; Yodis *et al.,* 2003). The annual average precipitation is about 1600 mm, with the lowest monthly average of 86 mm in October and the highest monthly average of 159 mm in July (Wu, 2005).

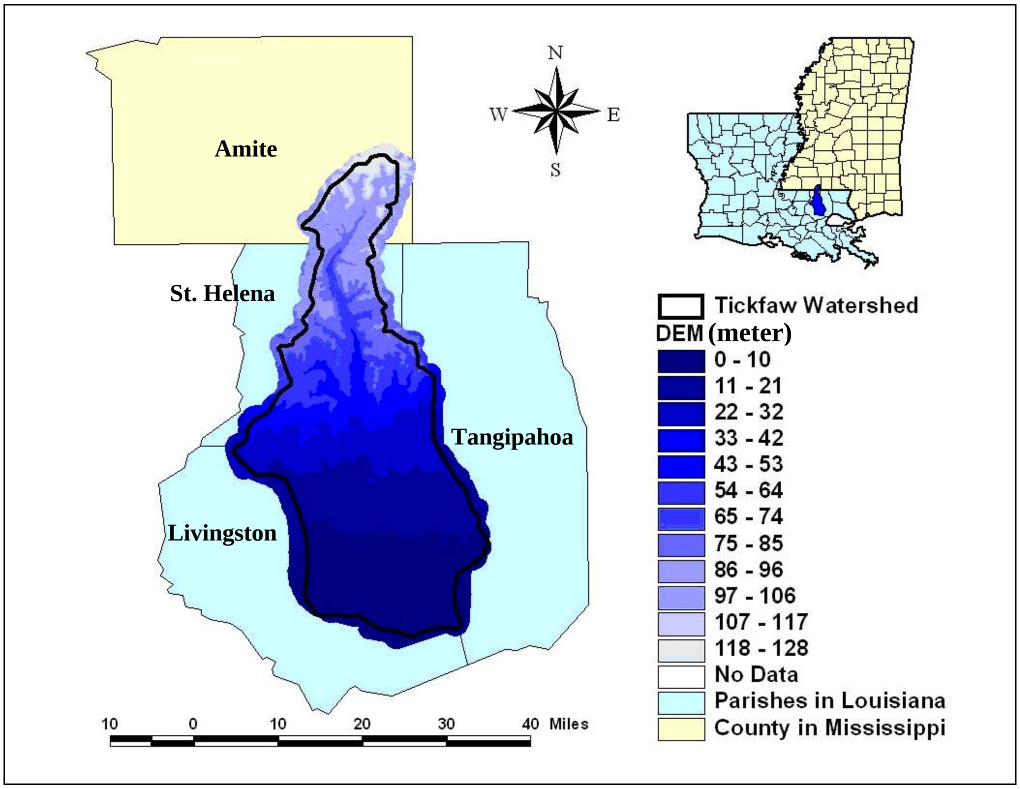


Figure 3.1. The location of study area – Tickfaw River watershed

The major part of Tickfaw watershed is covered by hills and terraces (Penland et al, 2002). The soil types in the study area include Pleistocene Terrace soils, Flatwoods soils, and Loess soils (Johnson and Yodis, 1998; Yodis *et al.,* 2003). The Pleistocene Terrace soils are dense, solid, and relatively impermeable layers. Flatwoods soils have high acidity and low fertility, and they are rarely exploited for crop agriculture. Loess soils have developed along the margins of the Mississippi River valley where loess deposits on the terrace complexes are of sufficient thickness for soil profile development (Johnson and Yodis, 1998; Yodis *et al.,* 2003). In the study area, longleaf pine and hardwoods are most common. The forests in the northern Tickfaw watershed are dominated by longleaf pine, and the forests in the southern Tickfaw watershed are dominated by the mixed forests of longleaf pine and hardwoods.

Tickfaw watershed is one of tributaries to Lake Pontchartrain. Based on a set of long-term records analysis done by Wu (2005), its mean daily discharge is 10.9 m^3/s. The major types of water bodies in Tickfaw watershed are rivers that include the Natalbany River, Pontchatoula River, Tickfaw River, and Yellow Water River. These rivers have been partially impaired by dissolved oxygen, fecal coliform, phosphorus, mercury, dissolved solid, lead, and nitrite/nitrate (Table 3.1). The major pollution sources are atmospheric deposition toxics, on-site treatment system, draining / filling /wetland loss, upstream source, drought-related impact, residential districts, and land development / redevelopment. Also a large portion of pollution source is unknown (U.S. EPA, 2002).

According to the U.S. Census Bureau (2006), the Tangipahoa, St. Helena, and Livingston Parishes experienced a 12.5%, 2.2%, and 25% increase in its population respectively from 2000 to 2006 (Table 3.2). Since the 1980s, this area has experienced a rapid urban growth and significant deforestation (Couvillion, 2005). Based on the census of three parishes and one county that are partially covered by Tickfaw watershed, the total population increased from

Table 3.1. Major pollutants and pollution sources in the Tickfaw River watershed (Source: U.S. EPA, 2002)

Pollutants	**Rivers, Streams, Creeks (Miles)**		**Pollution Sources**	**Rivers, Streams, Creeks (Miles)**
Chloride	26		Atmospheric Deposition Toxics	173
Lead	25		Drainage / Filling / Wetland Loss	94
Mercury	173		Drought-Related Impacts	38
Oxygen Dissolved	25		On-site Treatment System / Similar System	159
Sulfates	26		Residential Districts	25
Total Dissolved Solid	131		Site Clearance	12
Total Fecal Coliform	159		Upstream Source	173
			Source Unknown	68

Table 3.2. The population change in the study area
(Source: U.S. Census Bureau, 2006)

Parishes or Counties	Land Area ($mile^2$)	Population		Population Change	Population Density in 2000 (Persons/ $mile^2$)	Households In 2000
		2000	2006			
Tangipahoa	790.24	100,588	113,137	12.5%	137.3	36,558
St. Helena	408.36	10,525	10,759	2.2%	25.8	3,873
Livingston	648.02	91,814	114,805	25.0%	141.7	32,630
Amite	729.60	13,599	13,366	-1.0%	18.6	5,271

179,237 to 239,566 during 1990-2005. The increasing rate is almost 34%. These increases are related to urbanization and economic development. The major economic activities in the study area are forestry and cattle / poultry production (Johnson and Yodis, 1998; Yodis *et al*., 2003). Both forestry and cattle production are major part of the gross income in Tangipahoa, Helena, and Livingston Parishes. Poultry production is most important in both Helena and Tickfaw Parishes. The secondary economic activities include wood production and gravel mining. There are, totally, about twenty sawmills and millwork / cabinetry sites in those three parishes. Also, a significant amount of historic and active sand /gravel mines are identified within the watershed. The watershed of Tickfaw River has been extensively mined with little or no reclamation (Lake Pontchartrain Basin Foundation, 2005).

3.1.2. Data Sources

The primary data were a set of cloud-free time serial Landsat TM/ETM+ images. These images were acquired during 1986-2005. Six of them are Landsat 5 TM images acquired in 1986, 1990, 1992, 1995, 1997, and 2005, while the remaining two are Landsat 7 ETM+ images acquired in 2000 and 2002 (Figure 3.2, Figure 3.3). The characteristics of these images are displayed in Table 3.3. The preferred season for the Landsat images is in winter (January and February). When vegetation is in the stage of hibernation and clouds do not often occur, it is much easier to obtain time serial cloud-free image for the research. The major data sources include U.S. Geological Survey (USGS, 2005), Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, SD (EROS, 2005), and Global Land Cover Facility (GLCF) in University of Maryland, College Park, MD (GLCF, 2005). The Landsat 5 TM and Landsat 7 ETM+ images used in the study correspond to World Reference System 2 (WRS-2), Path 22 and Row 39. They have 28.5 m or 30 m of spatial resolution, 7 bands or 8 bands of spectral resolution, 8 bits of radiometric resolution, and 16 days of temporal resolution.

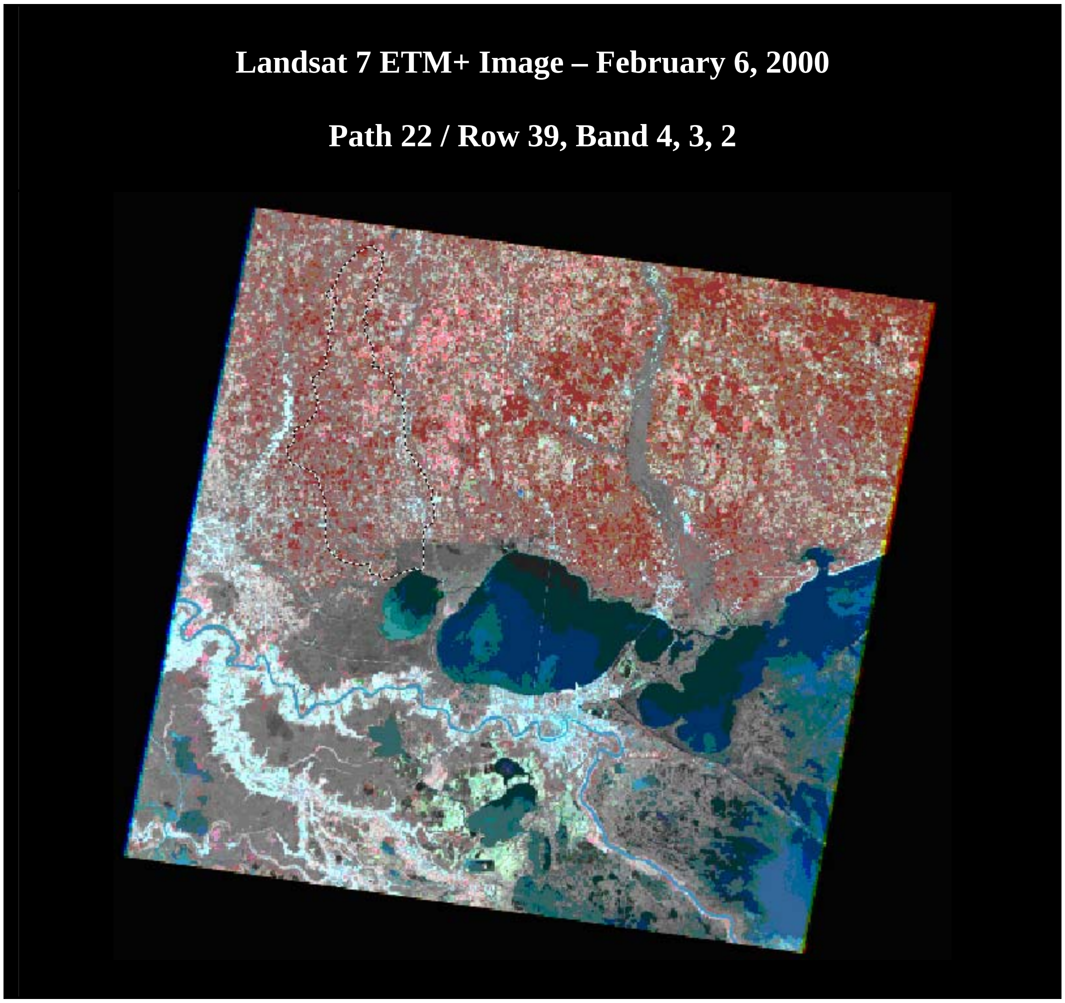


Figure 3.2. One of the primary images – Landsat 7 ETM+ Image
(Acquired on February 6, 2000, and Source: EROS, 2005)

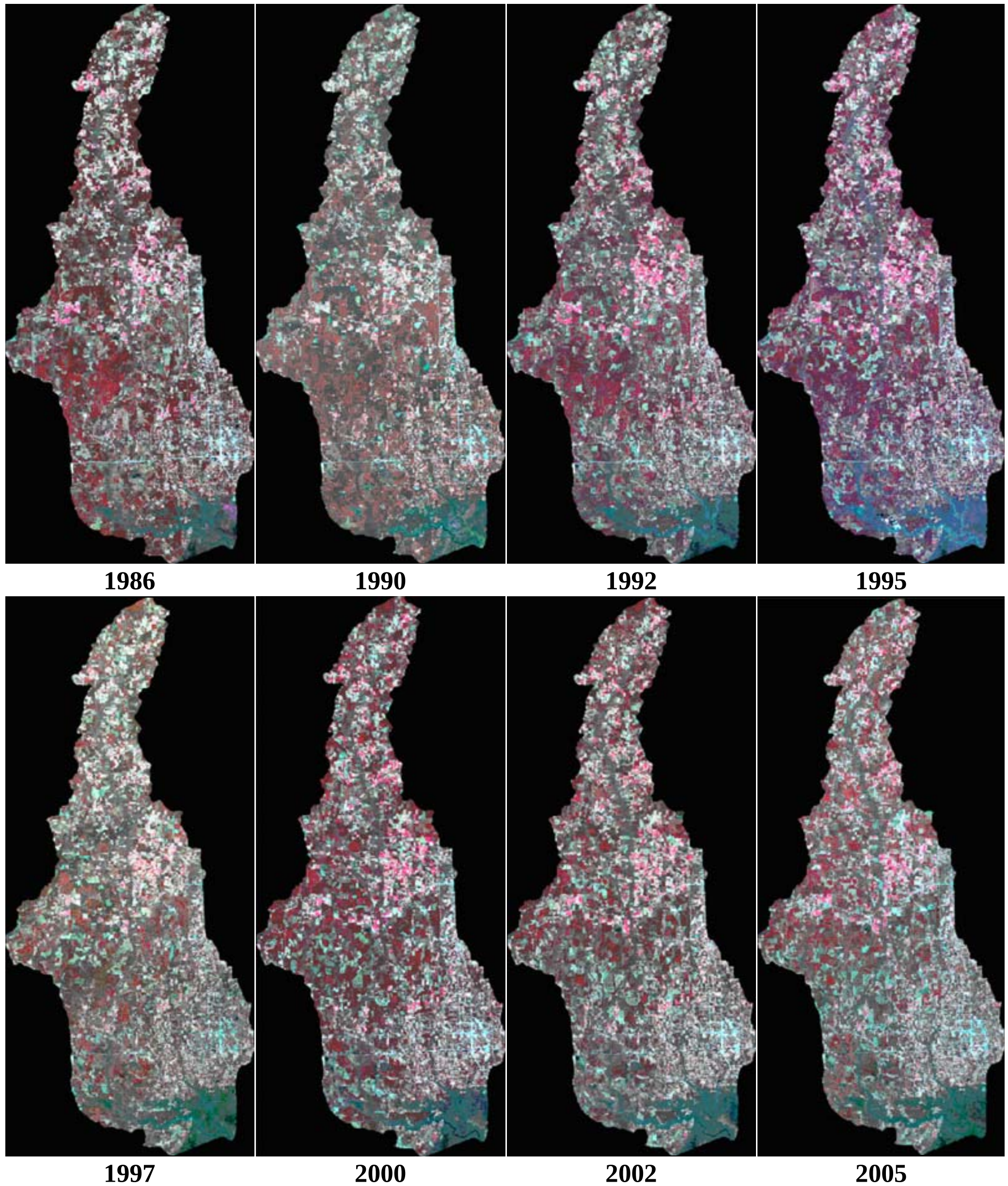


Figure 3.3. Landsat TM/ETM+ images of Tickfaw watershed
from 1986 to 2005 displayed in BGR for Band 4, 3, 2
(Sources: USGS, 2005; EROS, 2005; and GLCF, 2005)

Table 3.3. The characteristics of primary data geo-referencing, and radiometric correction

Image Entity ID	Satellite & Image	Acquired Date*	WRS 2 Path & Row	Cloud Cover (%)	Spatial Resolution (m)**	Sun Elevation (˚)	Julian Day	Geometric Correction RMSE	Radiometric Correction
5022039000504210	Landsat 5 TM Image	02/11/2005	22 / 39	0%	28.5	37.34645784	42	0.4160	yes
7022039000205850	Landsat 7 ETM+ Image	02/27/2002	22 / 39	0%	28.5	42.9555779	58	0.5183	yes
7022039000003750	Landsat 7 ETM+ Image	02/06/2000	22 / 39	0%	28.5	36.9569130	37	0.3264	yes
5022039009702010	Landsat 5 TM Image	01/20/1997	22 / 39	0%	28.5	30.03385664	20	0.4710	yes
5022039009530110	Landsat 5 TM Image	01/13/1995	22 / 39	0%	28.5	30.19635265	13	Reference	yes
5022039009203910	Landsat 5 TM Image	02/08/1992	22 / 39	0%	28.5	33.69316808	39	0.3584	yes
5022039009030510	Landsat 5 TM Image	11/01/1990	22 / 39	0%	30.0	37.64171782	305	0.4773	yes
5022039008605410	Landsat 5 TM Image	02/23/1986	22 / 39	0%	28.5	38.3904829	54	0.6153	yes
Data sources	* 2002 image was acquired from Mr. DeWitt Braud in LSU. 1990 image was downloaded from GLCF web site. Rest of images were purchased from EROS. ** Projection was changed into UTM, 15 zone, GRS 1980, Datum NAD83. Spatial resolution was changed into 30 m.								

The secondary data for the study included Digital Elevation Models (DEMs), Soil Data, Digital Orthophoto Quarter Quadrangles (DOQQs), and LCLU Data. Both DEMs and Soil data come with Better Assessment Science Integrating Point & Nonpoint Sources (BASINS) version 3.0, provided by United States Environmental Protection Agency (U.S. EPA, 2001). The scale of DEM data is 1:250,000 and data were converted into raster data with 30 meter resolution. Soil data contain soil category information, and data were converted into raster data based on the various soil type. DOQQs acquired in 1988 were purchased from EROS Data Center in Sioux Falls, SD. DOQQs acquired in 1998 and 2004 were free required from Computer Aided Design and Geographic Information System (CADGIS) Research Laboratory at Louisiana State University (CADGIS, 1984). These DOQQ images were used for accuracy assessment. U.S. National Land Cover Dataset 1992 and 2001 were obtained from Multi-Resolution Land Characteristics (MRLC) Consortium (MRLC, 1992; 2001). Coastal Louisiana land cover change data (1990-1996) were obtained from NOAA Coastal Services Center (NOAA, 2006). All DOQQ images and LCLU data were used for accuracy assessment only.

3.2. Data Preprocessing

When carrying out an absolute comparison between various dates or periods are to be carried out, the data preprocessing of multi-temporal remotely sensed imagery is very important and demanding (Coppin *et al.,* 2004). A successful implementation of LCLU study using remotely-sensed images requires careful considerations of the remote sensing system, environmental characteristics, and image processing procedures. The spectral, spatial, temporal, and radiometric resolutions of remote sensing images have significant impact on the success of a LCLU study (Lu *et al*., 2004a). Of the different aspects of data pre-processing for time serial images-based spatial analysis and modeling, geometric rectification or images registration, radiometric correction, and atmospheric correction are the most important. Spatial information, transformation and indexing information derived from the remote sensing image, and ancillary data can also improve LCLU study (Huang and Jenson, 1997; Lu *et al*., 2004a).

3.2.1. Geometric Correction and Radiometric Correction

Geometric correction is a procedure to correct the systematic and nonsystematic errors in the remote sensing system during the process of image acquisition. It involves a mathematical transformation of the coordinates. In this research, an image-to-image rectification-based geometric correction was carried out between the reference image and rest of the Landsat TM/ETM+ images respectively. A Landsat TM 5 image, acquired on January 13, 1995, was used as a reference for geometric correction purposes. The other Landsat TM/ETM+ images were registered to the reference data. During the geometric correction, 10 ground control points (GCPs), 20 check points, a third-order polynomial transformation, and nearest neighbor re-sampling were selected. All Landsat TM/ETM+ images were re-sampled to 30m and re-projected to UTM Zone 15 North with spheroid of GRS 1980, datum of NAD 83 North. The rectification with RMSE error of less than 1 pixel was required.

During data acquisition, variations in solar illumination condition, such as the solar elevation, Julian day, and time of day (Table 3.3), will cause significant radiometric shift among images (Forster, 1984; Hall *et al*., 1991). Additionally, the variations in atmospheric scattering / absorption and in detector performance will also cause significant radiometric differences. Therefore, the time serial images should be corrected radiometrically. Radiometric correction is a procedure to correct the errors caused by those factors. In this research, a set of models accounting for many of such factors was provided by Mr. DeWitt Braud in the Coastal Studies Institute, LSU. The radiometric correction models took Julian date as input and output earth-sun distance. The earth-sun distance, sun elevation, and resultant images from geometric correction were taken as input data, and the output reflectance images were corrected radiometrically.

3.2.2. Transformation and Indexing

The Tasselled Cap Transformation (TCT) was derived by Kauth and Thomas in 1976 (Tso and Mather, 2001). TCT is a linear transformation of the six TM / ETM+ reflective bands that creates a six-band image with the following characteristics (Campbell, 1996):

Band 1 => Soil brightness index

Band 2 => Vegetation greenness index

Band 3 => Canopy and soil moisture wetness index

Band 4-6 => Atmospheric haze components

Usually, the brightness, greenness, and wetness are related to urban / bare land, forest / range / crop, and wetland / water respectively. The top three bands in the resultant image have very high percentage of variance explained, and they are used often in further image analysis and modeling.

Principal Components Analysis (PCA) is a multivariate analysis technique which can be used for compacting redundant spectral components into fewer principal components based on the most variance in the original image (Singh, 1989). It is usually used for compressing multi-dimension image data. In this research, each Landsat TM/ETM+ image was transformed using PCA, and output was a six-band image with 8 bits of radiometric resolution. The eigenvalues of principal components were used to determine which component will be selected for the further analysis. The first principal component always has the largest percent of variance explained, and in this research, the top three principal components usually contained most information. Therefore these are used for the next study procedure.

Texture is the frequency of tonal change within the image that arises when a number of small features are viewed together. It gives the visual impression of the roughness or smoothness of an object (Lo and Yeung, 2007). Shaban and Dikshit (2001) found that the combination of texture and spectral features improved the overall classification accuracy. Compared to the obtained result based solely on spectral features, about 9% and 17% increases were achieved for an addition of one and two textures, respectively. In this research, texture analysis was performed on each Landsat TM/ETM+ image by using Mean Euclidean Distance (MED) algorithm, and the resultant images were used in further analysis.

Iterative Self-Organizing Data Analysis (ISODATA) is commonly used clustering technique. It uses cluster analysis to produce natural clusters of pixels of similar brightness values from the multi-spectral image data (Lo and Yeung, 2007). It can be used as an information compacting technique also. In this research, ISODATA was performed on each Landsat TM/ETM+ image, and a set of ISODATA images with 50 classes were created.

Vegetation is one of the important land features in the study area, and it is important to distinguish vegetated area and non-vegetated area. The Normalized Difference Vegetation Index (NDVI) can be used to extract vegetation information from Landsat TM/ETM+ images. It is the normalization of {(Band4-Band3) / (Band4+Band3)} (Mas, 1999; Lyon *et al*., 1998). Similar to NDVI, a Normalized Land Water Index (NLWI), proposed by Mr. DeWitt Braud in LSU, was used to create a land water interface. It is the normalized band ratio of {(Band1+Band2+Band3) / (Band4+Band5+Band7)}. In this research, both NDVI and NLWI derived from each Landsat TM/ETM+ image are created and used for the further analysis.

3.2.3. Conversion and Layer Stack

DEM is the measurement of height above a datum and it is related to the absolute altitudes or elevations of the points contained in the data (Lo and Yeung, 2007). A 1:250,000 scale DEM was retrieved from the BASINS 3.0 package developed by the U.S. Environmental Protection Agency (U.S. EPA, 2001). It was a shape file, and the elevation ranges from 0 to 127 meters. Soil data was also retrieved from BASINS 3.0, and it was originally a State Soil

Geographic Data Set (STATSGO). The soil data set was a shape file containing 22 classes. To develop a layer stack for classification purposes, both DEM and soil were converted from vector into raster data. They were also re-sampled to 30m and re-projected to UTM Zone 15 North with spheroid of GRS 1980, datum of NAD 83 North.

3.3. Research Methods

The research involves three tasks: LCLU classification, change detection, and predictive modeling. The description for each task is given in the following sections.

3.3.1. Classification

In this research, based on the consideration on the categories of interest, data characteristics, study objectives, time limitation, and capability to access particular resources, only first-level classification in Anderson's classification system was used with minor modifications (Table 3.4).

Four classification methods were involved in this study: unsupervised classification, supervised classification, hybrid classification, and knowledge-based classification. The first three types of classification were performed on ERDAS Imagine 9.1, and the GA-based classification was performed on ERDAS Imagine 9.1 and GATree 2.0 (Papagelis and Kalles, 2000; 2001; Kalles and Papagelis, 2006). The comparison of different classification methods was carried out using 2000 Landsat 7 ETM+ images. The performance of these methods was compared based on the classification accuracy. With the consideration of temporal difference among DOQQs and the 2000 Landsat 7 ETM+ image, both the 1998 DOQQs and the 2004 DOQQs were simultaneously used for accuracy assessment. Stratified random sampling was used to select reference points, and minimum sampling size (21 points) for each class was set up based on the smallest class. During accuracy assessment, if the same sampling pixel on both DOQQs suggested a different feature, it is discarded, and a new sample pixel was selected until

Table 3.4. The land-cover and land-use classification scheme
(Source: modified from Anderson *et al*., 1976)

Land Feature Classes	Land Feature Description	Classification Code	Color
Urban	Residential, commercial, industrial, transportation, other urban land	1	
Agricultural Land	Cropland, pasture, orchards, grove, feeding operations, other agricultural land	2	
Grass / Shrub	Herbaceous rangeland, shrub and brush rangeland, mixed rangeland	3	
Forest	Deciduous forest, evergreen forest, mixed forest land.	4	
Water	Streams, canals, lakes, reservoirs, bays, and estuaries	5	
Wetland	Forested wetland, non-forested wetland	6	
Other	Sandy area, mining site, transitional area, mixed barren land,	7	

1,000 pixels were defined. The accuracy assessment method was also used for other Landsat TM/ETM+ images classification in this research.

3.3.2. Change Detection

After classifying a set of time serial images, a LCLU layer stack was created, then 7992 pixels were randomly sampled for process-oriented change detection. The minimum training sample size (100 points) was set up based on the size of the smallest LCLU class that is OTHER in most cases. The training data was used to create a log file that records each pixel's change process. The log file was input to the process mining software ProM 4.0 that contains the genetic Petri net function developed by Alves De Medeiros *et al.* (2007). The output was a set of Petri nets. Since the change detection is process oriented, it is necessary to ensure whether the process model can represent the "real world" well or not. One way to do so is to compare the projected image with the classified image. Therefore, in the phase of accuracy assessment, a classified image was used as reference data that was assumed correct, and the projected image derived from the process model was compared with the reference data. The same procedure was also used for the calibrating predictive model.

3.3.3. Predictive Modeling

Predictive modeling provides future scenarios of the LCLU pattern based on the model of change process. In this research, a geographic cellular automata (CA) was designed and implemented for LCLU predictive modeling. The key component of geographic CA is transition rules to be developed using the process models derived from Petri nets. The other components of geographic CA are cell space, states, neighborhood, and time step. Based on the data characteristics and research objectives in this study, those components were configured as listed in Table 3.5. Different configurations of geographic CA will lead to different accuracies of LCLU predictive modeling, so it is important to calibrate the predictive model. The purpose of

Table 3.5. Configurations of the geographic cellular automata

Spatial Resolution	Dimension	Cell Space	Cell States	Time Steps	Neighborhood	Transition Rules
7.5 m	3 D	378 * 813	2 or 7	2 - 3 or 4 - 5 years	von Neumann or Moore at size 3x3 or 7x7	Process models at different evolution levels
30 m		1507 * 3249				
120 m		6025 * 12993				

calibration is to develop a good model with high predictive modeling accuracy. During the phase of calibration, different sets of configuration were tested and evaluated.

3.4. Algorithms

3.4.1. Genetic Algorithms

The genetic algorithm (GA) used in this study was proposed by Holland (1975) and Goldberg (1989), then implemented and extended by Wall (1996), and visualized by Papagelis and Kalles (2000; 2001). Combining their work, the major idea of the GA-based decision tree is given as follows (Figure 3.4). There are different data structures, such as binary string, array, list, and tree, for representing the solution of the problem in GAs. In GAs, a genome is a solution for the problem. Genomes in the search space are usually represented by binary strings. However, such string-based genome representations are not well suited for representing the space of concept descriptions that are generally symbolic in nature and with various length and complexity (Papagelis and Kalles, 2000; 2001). Thus, tree representation is used to build a population of minimal binary decision trees. In this study, the population, representing the search space of problem solutions, consists of trees. Each tree represents a problem solution and it refers to a genome. Genomes with high fitness have better chance to be reproduced (Papagelis and Kalles, 2000; 2001; Wall, 1996).

For a given GENETIC ALGORITHM (***Generation, Gsize, Psize, Rrate, Crate, Mrate***), ***Generations*** is a function that updates the generations value during evolution, ***Gsize*** is the size of generations specifying the termination criterion, ***Psize*** is the size of population, ***Rrate*** is the REPLACE rate, ***Crate*** is the CROSSOVER rate, and ***Mrate*** is the MUTATION rate.

Based on the input data, the algorithm randomly INITIALIZEs a population ***iPopulation*** with size ***Psize***. In this case, the population is a set of decision trees that are commonly derived from Quinlan’s ID3 algorithm (Mitchell, 1997). Since ID3 is a popular algorithm for generating

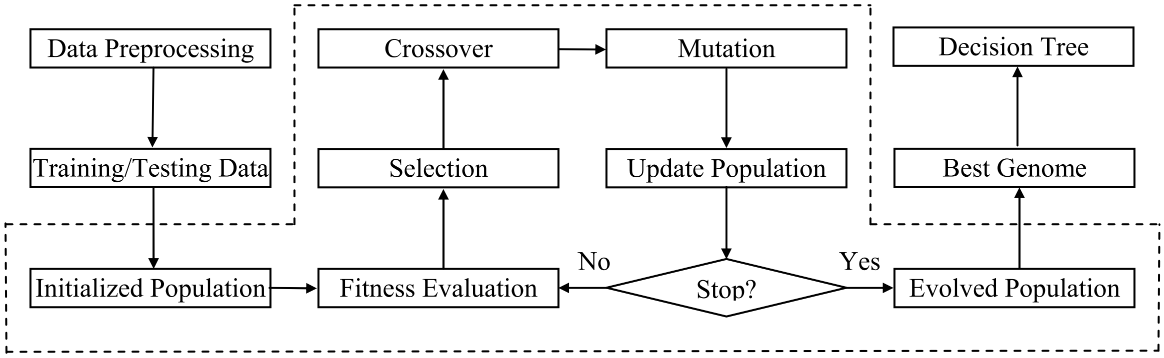


Figure 3.4. The workflow of genetic algorithms

adecision tree, the procedure of initializing a population or creating decision trees using ID3 algorithm will not be demonstrated here.

After initializing a population, the algorithm EVALUATEs the ***Fitness(treei)*** for each individual (or tree) in ***iPopulation***. The ***Fitness(treei)*** is calculated by the following equation (Papagelis and Kalles, 2000; 2001; Wall, 2006):

$$Fitness(tree_i) = CorrectClassified_i^2 * \frac{x}{size_i^2 + x}$$

here, ***CorrectClassifiedi*** is the overall classification accuracy when ***treei*** is used to classify training data; ***sizei*** is the number of leaves on the decision tree ***i***; ***x*** is an arbitrary large number. When the size of the tree is small, the size factor is near one. It decreases when the tree grows. In this way, the payoff is greater for smaller trees in terms of fitness value (Papagelis and Kalles, 2000; 2001).

After evaluating the initial population, the system goes into a WHILE DO loop, and each loop refers to a generation. First, based on ***Crate*** and ***Fitness(treei)***, ***Crate**** ***iPopulation*** members are probabilistically selected from the ***iPopulation*** and added to the new population ***nPopulation***. The probability of selecting individual (***hi***) is calculated by equation:

$$\Pr(hi) = \frac{Fitness(hi)}{\sum_{j=1}^{Psize} Fitness(hj)}$$

here, ***Pr(hi)*** is probability of selecting individual (***hi***) (Goldberg, 1989).

The major operators are crossover and mutation. CROSSOVER occurs between the selected (***Crate**** ***iPopulation*****)/2** pairs of individuals, each pair produces two offsprings, and those offsprings will be added to the ***nPopulation***. MUTATION will occur among the ***Mrate*****nPopulatio*** individuals, each individual produces one offsprings, and those offsprings will be added to ***nPopulation*** also.

Based on the generations gap ***Rrate***, ***Rrate*iPopulation*** individuals of initial population are replaced by individuals in new population ***nPopulation***. Then, ***Generations*= *Generations*+1** function update generations number**.** If ***Generations***is larger than ***Gsize***, the WHILE DO loop will be terminated and return an evolved population. Otherwise, it goes to next loop (or generation).

Figures 3.5 and 3.6 show a couple of examples on how genetic operators work. Crossover operator chooses two random nodes and swaps those nodes' sub-trees. Mutation operator chooses a random node and replaces the node's test-value with a new randomly chosen value. The mutation operator could also occur on leaves in which it replaces the installed class with a new random chosen class (Papagelis and Kalles, 2000; 2001). The examples in Figures 3.5 and 3.6 were used just to demonstrate how genetic operators work, and the decision trees developed in this study are much more complex than them. In GATree 2.0, the size of population or the number of trees was defined by the user. Each tree is a solution for the entire LCLU classification problem. It has one root, a number of nodes, and a number of leaves. Each leaf represents a class instance, such as URBAN, FOREST, or WATER. The number of leaves also refers to the size of tree. The tree that has the highest fitness is the best solution. When a small size of decision tree is used to classify LCLU, the tree with high fitness generally has higher classification accuracy, and high fitness decision tree has a high probability of being selected.

3.4.2. Petri Net

Petri nets (PNs) have been commonly used to extract the knowledge of processes from event logs in business and manufacturing field. According to the research done by van der Aalst and Weijters (2005), the assumption of PNs application is that it is possible to record events so that 1) each event refers to an activity, 2) each event refers to a case, 3) each event can have a performer, also referred to as originator, and 4) events have a time stamp and are totally ordered.

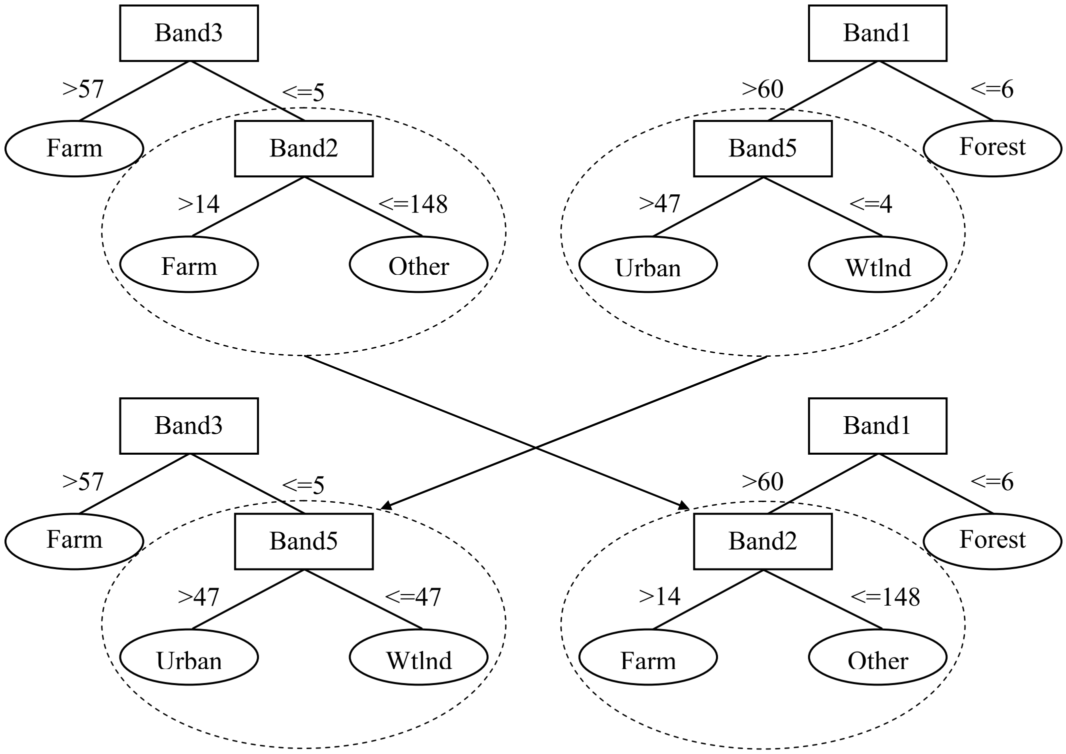


Figure 3.5. Crossover operation on sub-trees
(the values are used for describing the crossover operation)

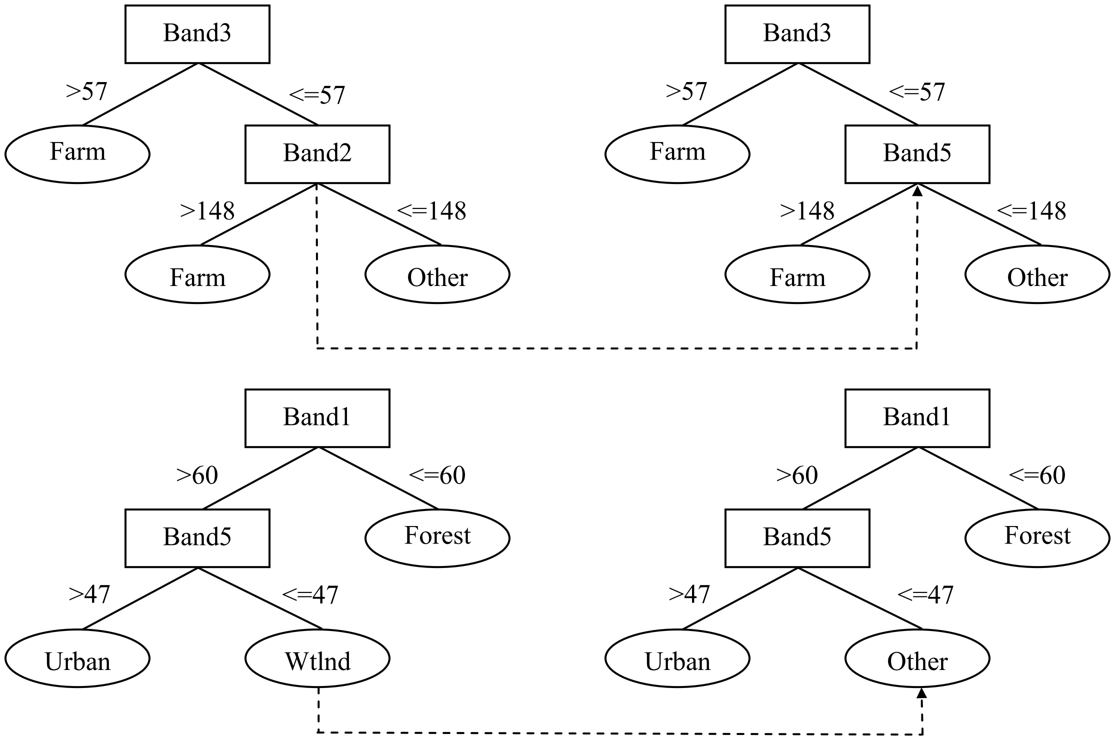


Figure 3.6. Mutation operation on sub-trees
(the values are used for describing the mutation operation)

In LCLU study, after accurately classifying a set of time serial Landsat TM/ETM+ images and creating a LCLU stack, each pixel's category is recorded. An event refers to a pixel holding a class value at a particular time. All ordered pixels (on different layers) with same coordinate and different time stamps consist of a process, each pixel holds a value that can be regarded as an event (Table 3.6). Because every event has a time stamp, a set of ordered pixels with same coordinate consists of a process instance. Therefore, a spatio-temporal process mining can be carried out. Process mining aims at extracting knowledge from a set of process instances.

The classic PNs, namely Place / Transition (P/T) nets, were used to introduce the basic concept of PN (David and Alla, 1992; van der Aalst *et al*., 2004). A PN has two types of nodes, namely places and transitions (Figure 3.7). A place is represented by a circle, and a transition by a box. Places and transitions are connected by arcs. The number of places and transitions is finite and not zero. An arc is directed and connects either a place to a transition or a transition to a place. In other words, a PN can be defined as a tuple (P, T, F) (Desel, 2005), where:

- P is a finite set of place represented by a circle, and it may hold token(s) represented by a black dot,
- T is a finite set of transitions represented by a square, and it may contain different operators (e.g., AND-split, AND-join),
- $F \subseteq \left(P \times T \right) \cup \left(T \times P \right)$ is a set of directed arcs or flow relation, and it connects place and transition.

A marked *P/T* net is a pair (*N, s*), where $N = (P, T, F)$ is a *P/T* net and where ***s*** is a bag over P denoting the marked of the net, i.e. $s \in P \rightarrow IN$. The set of all marked *P/T* nets is denoted *N*.

A *P/T* net consists of places, transitions, token. Place includes input place and output place, transition has AND-split and AND-join transition, and a black dot represents a token. The dynamic behavior of such a marked P/T-net is defined by a firing rule. A PN can be used to

Table 3.6. A land-cover and land-use change log
(Source: developed based on van der Aalst and Weijters, 2005)

Pixel ID	Process ID	Year1986	Year1990	Year1995	Year2000
Pixel 1	Process 1	FOREST	GRASS	FARM	URBAN
Pixel 2	Process 2	FOREST	FARM	GRASS	URBAN
Pixel 3	Process 3	FOREST	GRASS	FARM	URBAN
Pixel 4	Process 4	FOREST	FARM	GRASS	URBAN
Pixel 5	Process 5	FOREST	OTHER	--	URBAN

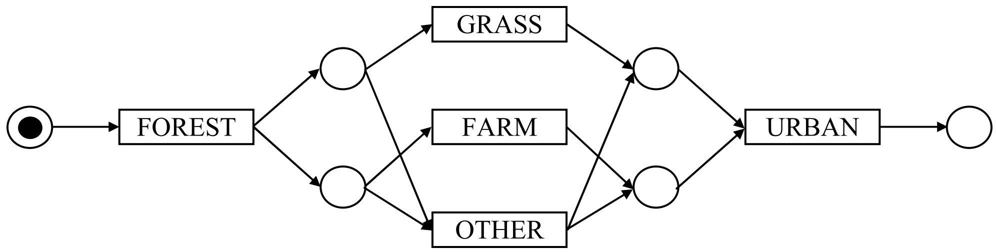


Figure 3.7. A process model corresponding to the event log shown in Table 3.6
(Source: developed based on van der Aalst and Weijters, 2005)

specify the routing of events. Events are modeled by transitions and causal dependencies are modeled by places and arcs. In fact, a place corresponds to a condition that can be used as pre-condition and / or post-condition for events. The LCLU change log is a set of event traces. Each trace corresponds to an execution of a process. The same process may occur many times in a log. In this situation, a different process followed the same path (P, T, F) (Desel, 2005).

Now, the question is how to extract those paths from an event log file? Many algorithms have been developed since 1998 (Cook and Wolf, 1998; Pinter and Golani, 2004; Herbst and Karagiannis, 2004; Schimm, 2004; and van der Aalst, 2003). Among them, α-Algorithm proposed by van der Aalst *et al*. (2005) is the most successful one for discovering processes from event log file. With the consideration of LCLU change study, the basic concept of α-Algorithm is introduced by using the simple example data in Table 3.6. Let ***L*** be a LCLU change log over ***T***, the process mining algorithm (α (L)) builds a net (P_L, T_L, F_L) through following steps (van der Aalst *et al*., 2003, 2004, 2005; van der Aalst and Weijters, 2005):

1) $T_L = \{t \in T \mid \exists_{\sigma \in L} t \in \sigma\}$,
2) $T_I = \{t \in T \mid \exists_{\sigma \in L} t \in first(\sigma)\}$,
3) $T_O = L\{t \in T \mid \exists_{\sigma \in L} t \in last(\sigma)\}$,
4) $X_L = \{(A,B) \mid A \subseteq T_L \wedge B \subseteq T_L$
$\wedge \forall_{a \in A} \forall_{b \in B} a \rightarrow_L b \wedge \forall_{a1, a2 \in A} a1 \#_L a2$
$\wedge \forall_{b1, b2 \in B} b1 \#_L b2\}$,
5) $Y_L = \{(A,B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A', B'')$,
6) $P_L = \{p_{(A,B)} \mid (A,B) \in Y_L\} \bigcup \{i_L, o_L\}$,
7) $F_L = \{(a, p_{(A,B)}) \mid (A,B) \in (Y_L \wedge a \in A\}$
$\bigcup \{p_{(A,B)}, b) \mid (A,B) \in (Y_L \wedge b \in B\}$
$\bigcup \{(i_L, t) \mid (t \in T_I) \bigcup \{(t, o_L) \mid (t \in T_o), and$
8) $\alpha(L) = (P_L, T_L, F_L)$.

Step 1 – the algorithm examines the LCLU change log (Table 3.6) and identifies all events appearing in the log. In other words, it creates a set of transition (***TL***), and $\boldsymbol{T_L}$ = *(FOREST, GRASS, FARM, OTHER, URBAN)*.

Step 2 – the algorithm examines the LCLU change log and identifies the set of initial (or input) events by simply checking the earliest events in the LCLU change log. For example, in this case, the earliest year is 1986 (Table 3.6). It creates a set of initial (or input) transition (T_I), and T_I = *(FOREST)*.

Step 3 – the algorithm examines the LCLU change log and identifies the set of final (or output) events by simply checking the latest events in the LCLU change log. For example, in this case, the latest year is 2000 (Table 3.6). It creates a set of final (or output) transition (T_O), and T_O = *(URBAN)*.

Step 4 – the algorithm identifies all causally related transition or relationship (X_L) among the events in the LCLU change log, and X_L = *{({ FOREST }, { GRASS }), ({ FOREST }, { FARM }), ({ FOREST }, { OTHER }), ({ GRASS }, { URBAN }), ({ FARM }, { URBAN }), ({ OTHER }, { URBAN }),({ FOREST }, { GRASS, OTHER }), ({ FOREST }, { FARM, OTHER }), ({ GRASS, OTHER }, { URBAN }), ({ FARM, OTHER }, { URBAN })}*.

Step 5 – the algorithm constructs only the minimal causality relationship (Y_L) based on (X_L) by taking only the largest elements with respect to set inclusion, and Y_L = *{({ FOREST }, { GRASS, OTHER }), ({ FOREST }, { FARM, OTHER }), ({ GRASS, OTHER }, { URBAN }), ({ FARM, OTHER }, { URBAN })}*.

Step 6 – the algorithm creates a set of places in the resultant Petri net. For example, *P(FOREST, GRASS)* is a place connecting transitions in FOREST with transition in GRASS, i_L is the unique input place denoting the start of the process, and o_L is the unique output place denoting the end of the process. P_L = *{* i_L, o_L, *P({ FOREST }, { GRASS, OTHER }), P({ FOREST }, { FARM, OTHER }), P({ GRASS, OTHER }, { URBAN }), P({ FARM, OTHER }, { URBAN })}*.

Step 7 – the algorithm creates a set of connecting arcs in the resultant Petri net. For example, *(FOREST, GRASS)* is an arc connecting transitions in FOREST with transition in

GRASS. $F_L = \{(i_L, FOREST), (FOREST, P(\{ FOREST \}, \{ GRASS, OTHER \})), (P(\{ FOREST \}, \{ GRASS, OTHER \}), GRASS) \ldots\ldots (URBAN, o_L)\}$.

Step 8 – finally, the algorithm returns the discovered process model represented by a Petri net with place P_L, transition T_L, and arcs F_L, namely, $\partial (L) = (P_L, T_L, F_L)$ (Figure 3.8).

The next question is how the algorithm constructs a Petri net. Based on the α-Algorithm proposed by van der Aalst and Weijters (2005) and the LCLU change log (Table 3.6), a simple description is provided as follows. The algorithm assumes that two events, for example FOREST and GRASS, are connected through places if and only if $FOREST \rightarrow_L GRASS$ (Figure 3.8a). After event FOREST, if two events (for example GRASS and OTHER) are concurrent, they can occur in any order. Therefore, the algorithm assumes event GRASS and OTHER concurrent if and only if $GRASS \parallel_L OTHER$ (Figure 3.8b). This is the case of AND-split relation, and its counterpart AND-join is shown in Figure 3.8d. If those two events are not concurrent and they never follow each other directly, every time only one event happens. Therefore, the algorithm assumes event GRASS and OTHER are not concurrent and never follow each other if and only if $GRASS \#_L OTHER$ (Figure 3.8c). This is the case of XOR-split relation, and its counterpart XOR-join is shown in Figure 3.8e. Those relationships are represented by the foundation of α-Algorithm.

Although α-Algorithm has been successfully used to mine processes from an event log file, it still has some limitation. When an event log file contains incomplete, noisy, or rare instances, the algorithm is not able to extract processes. In order to overcome those limitations, genetic algorithm was introduced into the procedure of α-Algorithm-based process mining (Alves de Medeiros *et al*., 2007; van der Aalst *et al*., 2005; Schwardy, 2003; Weijters and Paredis, 2002). The major idea is that genetic operation (e.g., crossover, mutation, and fitness-based selection) can enlarge the search space and optimize the results. Theoretically, the

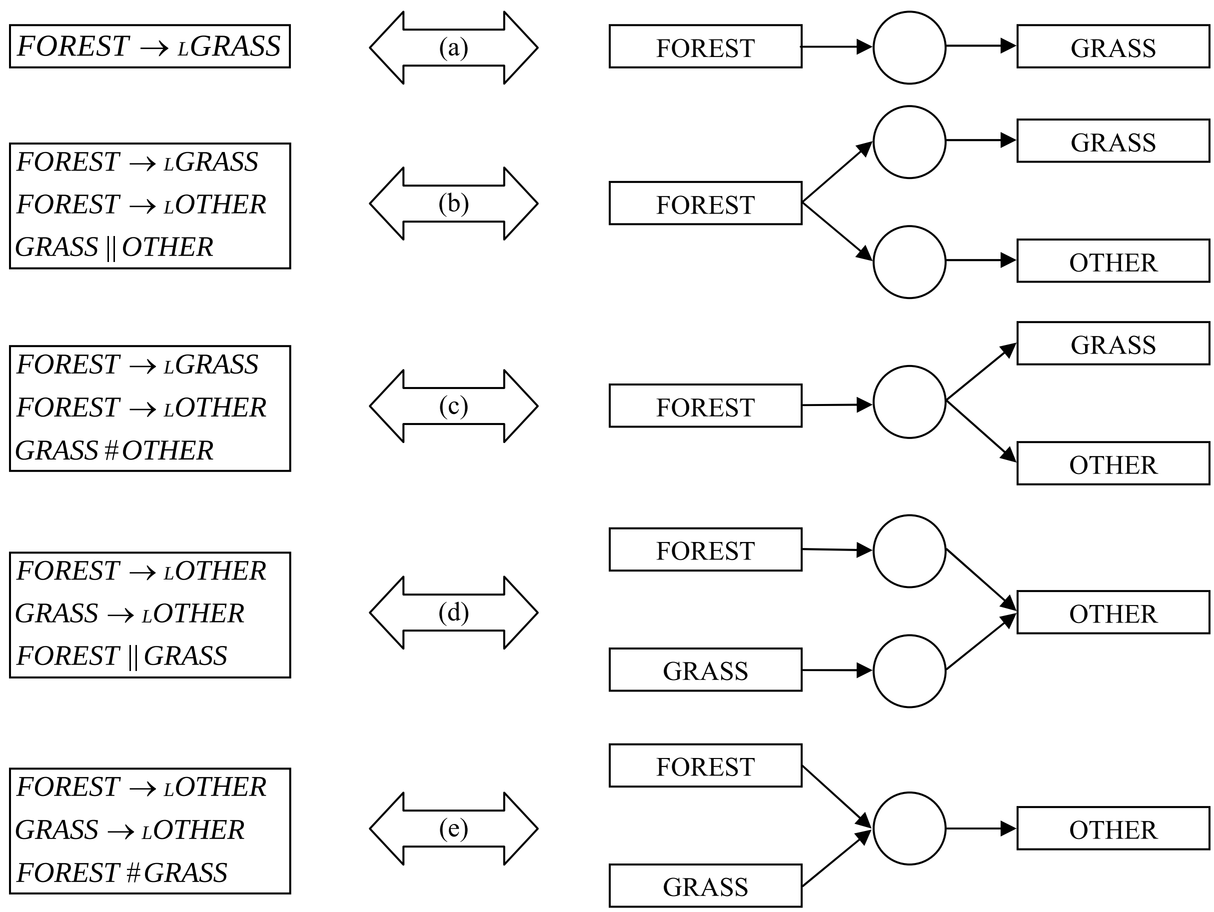


Figure 3.8. Constructing a Petri net using the log-based relationships $>_L$, $\rightarrow_L$, $\|_L$, and $\#_L$
(Source: developed based on van der Aalst and Weijters, 2005)

optimization is global because the search is based on an enlarged space. Figure 3.9 describes the main steps in GA-based PN for process mining. The major idea is the same as the one introduced in the previous section. But, the way in which genetic operators work and the way of calculating fitness are slightly different.

In GA-based PN, the individual is not a string or a decision tree; it is a process model represented by Petri net. An individual Petri net has place node, transition node, and directed arcs. Many individual PNs consist of a population, and genetic operations occur among various individuals. The most important and complex genetic operation in the GAs is the crossover operation. It aims at recombining existing materials in the current population. The starting point of the crossover operation is the two parents PN (e.g., parent PN #1 and parent PN #2). The results of applying the crossover are two offsprings (offspring PN #1 and offspring PN #2) (Alves de Medeiros *et al*., 2007; van der Aalst *et al*., 2005). When a crossover occurs, two parent PNs exchange genetic material (place nodes, transition nodes, and arcs), and the swap points at the parent PNs are randomly selected (Alves de Medeiros *et al*., 2007; van der Aalst *et al*., 2005).

Another important genetic operator is mutation. It aims at inserting new material in the current population, meaning that mutation operator may change the existing causality relations of a population. In GA-based PN, the mutation operator carries out one of following actions in an individual PN: 1) randomly select a PN and assign a task to it, or 2) randomly select a PN and remove a task from it.

The goal of process mining is to discover a process model from an event log. This mined process model should give a good insight about the behavior in the log. In other words, if the mined process model is complete, and it can reproduce (or parse) most process instances in the log. A process model should also be precise because a process model may parse extra processes if the mined process model is complete, and it can reproduce (or parse) most process instances in

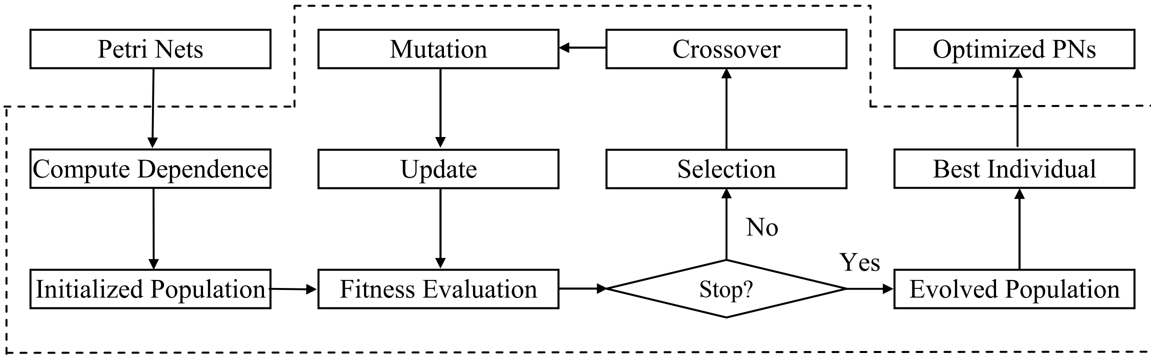


Figure 3.9. The workflow of genetic algorithms for optimizing process models

the log. A process model should also be precise because a process model may parse extra processes if instances do not belong to the log (Alves de Medeiros *et al*., 2007; van der Aalst *et al*., 2005). The fitness function guides the search process of the GA. The fitness of an individual PN is assessed by benefiting the individuals that can parse more event traces in the log and by punishing the individuals that allow for more extra behavior than the one expressed in the log. The fitness is strongly related to the number of correctly parsed traces from the event log. If an individual in the genetic population correctly describes the registered behavior in the event log, the fitness of that individual will be high. In this study, the fitness based evaluation of PN is similar to the approach that is described in the GA-based LCLU classification.

3.4.3. Cellular Automata

Since John von Neumann and Stanislaw Ulam proposed the concept of cellular automata (CA) in the 1950s, the original restrictive definition of CA has been extended to many different applications. In general, a CA is specified by the following definition (Deutsch and Dormann, 2005):

- Space – a regular discrete lattice of cells and boundary conditions.
- States – a finite and typically small set of values that characterize the cells.
- Neighborhood – a finite set of cells that surround an interested cell.
- Time steps – a finite set of time intervals that define the time to update all cells simultaneously, and
- Transition rules – a set of rules that determine the dynamics of cells' states.

Cellular automata are non-linear mathematical dynamic systems based on discrete space, time, and state (Deutsch and Dormann, 2005). A cellular automaton evolves in discrete time steps by updating its state according to the universal rule applied to each cell synchronously at each time step (Wolfram, 1984). The state of each cell is determined by both transition rules and neighborhood. Updated states of each cell are then the inputs for the next iteration. As iterations proceed, an initial cell configured with initial state of each cell evolved into a new and complex pattern. Usually, the behaviors of CA are complex. The idea of CA is that a simple transition rule can generate very complex evolving patterns and the evolving processes as a whole are very useful for aiding researchers in discovering underlying mechanisms of a real world system. As a type of no traditional mathematical models, the basic CA defined by Ulam, Von Neumann, and Conway is an abstract, simplified version of the real world and can be used to study the real world problems. However, it is not well suited for the applications in LCLU predictive modeling. Therefore, it is necessary to modify it from its formal characteristics as following:

- Cell space: the cell space is composed of individual cell. It usually refers to the number of the cells, the size of the cells, the shape of the cells, and the dimension of the CA. Although the majority of research has been based on 1-D or 2-D CA, 3-D CA have attracted attention in the literature since the early 1990s (Bays, 1990; 1991). Those characteristics of cell space are closely related to the study problem and data property. Because of the direct analogy of raster GIS and LCLU stack, only 3-D CA are considered in this dissertation research. The cells' number was same as the pixels' number in the AOI, the cells' shape was same as the square shape of the pixels, and the cells' size was same as the image spatial resolution.
- Cell states: the states of CA can be fixed or un-fixed. Some cells may never change their state, and other cells may change their states frequently. In a raster GIS based study the cell state can range from 0 to 255. In this dissertation research, the cell may hold any one of following states: urban, agricultural land, range, forest, water, wetland, and other land. Cells representing water or national forest park may never change its state, which means that its state is fixed. But cell representing other land

features may change their states more frequently, it means that its state is un-fixed.

- Neighborhood: a neighborhood is defined as a set of cells surrounding an interested cell (Figure 3.10). Traditional CA has two types of neighborhood: von Neumann and Moore. Each type of neighborhood has different sizes, such as 3 x 3, 5 x 5, and 7 x 7. The assumption is that just adjacent neighborhood cells have an influence on the interested cell. But the reality is that this assumption is not always true in the real world. In this research, both von Neumann and Moore neighborhood with different sizes (e.g. 3 x 3 and 7 x 7) were applied.
- Time steps: a geographic CA without a time step can only perform spatial modeling, but not spatio-temporal modeling. Time step is the cycle in which cells update their states simultaneously. In a LCLU predictive modeling study, the time steps of cells may be different and a group of cells may share a same time step. The time step in the process model is supposed to be relatively stable. In this research, each LCLU layer was temporally stamped, and the time step of the process model is 2-3 and 4–5 years.
- Transition rules: CA's transition rules define the behavior of cells. The rules should be explicit, and they can be developed based on the process models. In this research, a log file recording the behavior of sampling pixels, the α-Algorithm-based Petri nets were used to develop the process models, and genetic algorithms were used to optimize the process models. Finally, those process models were translated into explicit transition rules.

The configuration of geographic CA in this research is listed in Table 3.5. The combinations of different configurations were tested and the optimized geographic CA were used for LCLU predictive modeling. More detailed information about LCLU predictive modeling will be provided in Chapter 8.

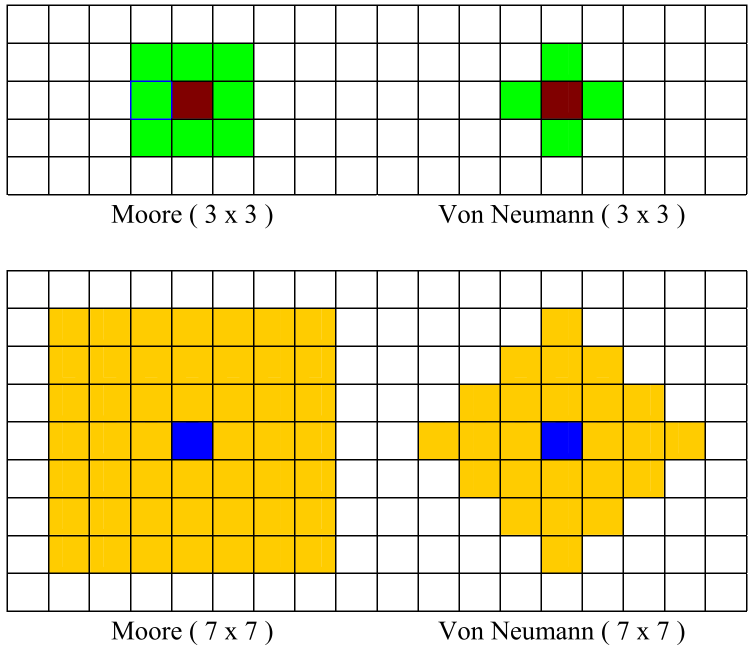


Figure 3.10. Two types of neighborhoods used in this research

3.5. Implementation

The major data processing and analyses were performed using a DELL desktop B110 (CPU2.2 GHZ, RAM 512 MB, HD 160), a DELL laptop B130 (CPU1.4, RAM 256, HD 40), and an External HD with 250 GB. Some works were done in different labs on the LSU campus. The data preprocessing includes geometric correction, radiometric correction, transformation, indexing, conversion, and layer stack. These procedures were performed in the environment of ArcGIS 9.0, ArcView 3.3, ERDAS Imagine 9.0, and Microsoft Excel.

Two data mining software packages were used in this dissertation study for LCLU classification purposes. The first one is GATree 2.0, and the second one is Weka 3. 5. Both of them contain genetic algorithms. In 1996, Wall implemented and extended the basic genetic algorithm by developing an open source C++ library of genetic algorithms – GAlib 2.4.6 (Wall, 1996). This library includes tools for using GAs to solve global search and optimization problems. By using GAlib, a visualized genetic algorithm based machine learning system, namely GATree 2.0, was developed by Papagelis and Kalles (2000; 2001). It was purchased and used to discover spatial / spectral knowledge in this study (www.gatree.com). Weka 3.5 is a different data mining software (Witten and Frank, 2005), which is developed using Java programming language (http://www.cs.waikato.ac.nz). Weka 3.5 supports many data mining algorithms including genetic algorithms, and it has to be run in the Java environment. It means that a Java JDK 1.5 should be installed first (http://www.java.com). Both of them were used in this research for spatio-spectral data mining and knowledge discovery.

Three other software packages were used in this research for process-oriented LCLU change detection, specifically for discovering spatio-temporal spectral knowledge. The first is an XML parser that can be used for parsing XML data files. The second is ProMimport 2.3 which takes input data file, and the third is ProM 4.0 which processes data file and output workflow or process models. All can be downloaded from http://is.tm.tue.nl. The resultant process models were optimized by adjusting the characteristics of data, the configuration of GAs, and the rate of pruning PNs.

For remote sensing data analysis and spatio-temporal modeling, ERDAS Imagine was used. A built-in cellular automaton in ERDAS Imagine should be the best way to perform geographic spatio-temporal modeling. In this study, the CA-based LCLU predictive model was created using the knowledge engineer in ERDAS Imagine 9.0. Also, the major components, such as cell space, cell states, neighborhood, time steps, and transition rules were built and encoded using ERDAS Imagine 9.0.

3.7. Summary

This chapter discussed the major characteristics of the study area, data sources, data preprocessing, and major research approaches (e.g., genetic algorithms, Petri nets, and cellular automata). The following summarizes the major points in the chapter:

- The study area, namely Tickfaw River watershed, is located in the southeastern Louisiana and the southwestern Mississippi, between 30˚19’14” - 31˚9’42” N latitude and 90˚31’58” - 90˚50’15” W longitude (Figure 3.1). It totally covers 771 mile^2 and includes a portion of Amite County, St. Helena Parish, Livingston Parish, and Tangipahoa Parish. The procedures of data preprocessing include geometric correction, radiometric correction, transformation, indexing, conversion, and layer stacking.
- A knowledge-based expert classifier was developed in this dissertation. GAs were used as a major spatial-spectral data mining and knowledge discovery technique. They include six parameters: the number of generation, the size of population, crossover rate, mutation rate, and generation gap. Those parameters can affect the

performance of GAs. The implementation of GAs-based knowledge discovery was based on GATree 2.0 and Weka 3.5. Both of them support complete GAs.

- A processes-oriented LCLU change detection approach was developed using both PNs and GAs. PNs were used as a major spatio-temporal spectral data mining and processes mining technique in this study. A PN includes state node, space node, arcs, and tokens. A traditional PN is not optimized. GAs were used to develop optimized PNs. The implementation of PNs and GAs-based PNs for LCLU change process mining was based on ProM 4.0 and ProMimport 2.3. They were used to take XML log data and output process models.
- A CA-based LCLU simulation model was developed for predicting future and reconstructing past LCLU pattern. A geographic cellular automaton consists of five components: cell space, cell states, neighborhood, time steps, and transition rules. The transition rule is the most important component of CA. In this research, a set of transition rules were developed from the LCLU change process models derived from PN, and they were encoded into a geographic cellular automaton using ERDAS Imagine. Because the process models were optimized using GAs, those transition rules can drive the CA efficiently.

CHAPTER 4 IMPACT OF GENETIC PARAMETERS ON LAND-COVER AND LAND-USE CLASSIFICATION ACCURACY

4.1. Experiment #1 Design

Figure 4.1 shows the procedures of applying GAs and expert system in LCLU classification. The Landsat 7 ETM image acquired on February 6, 2000 was used in this experiment. Bands 1 – 5 and Band 7 were selected. In order to take training samples from each band at same location, an ISODATA having 50 classes was created through ERDAS Imagine 9.0. The ISODATA was opened from the accuracy assessment window that was linked to the viewer containing the AOI of the Landsat 7 ETM image. A set of random samples (training / testing data) was created and classified. The category of each pixel was recorded. Both the coordinate and category of each pixel was copied to an Excel worksheet. An ASCII file that contains the coordinates for all random pixels was created based on the Excel worksheet. This coordinate file was then used to collect pixels' value from Bands 1 – 5 and Band 7 (Figure 4.2a). The size of the training data is 10000, and the size of testing data is 5000. After finishing random sampling, the data are converted into an ARFF file for input (Figure 4.2b). This file format is supported by most data mining / machine learning software packages including GATree 2.0 and Weka 3.5.

Based on the proposed questions related to research objectives, the following five genetic parameters were tested in this experiment: the number of generation, population size, crossover rate, mutation rate, and generation gap. For each test, one parameter's value is changed and tested, and the rest of the parameters remain at default. Each test generated a set of genomes. The best genome was represented by a decision tree, and it was translated into a set of classification rules (Table 4.1). The classification rules were then implemented to encode into an expert system using ERDAS Imagine 9.0.

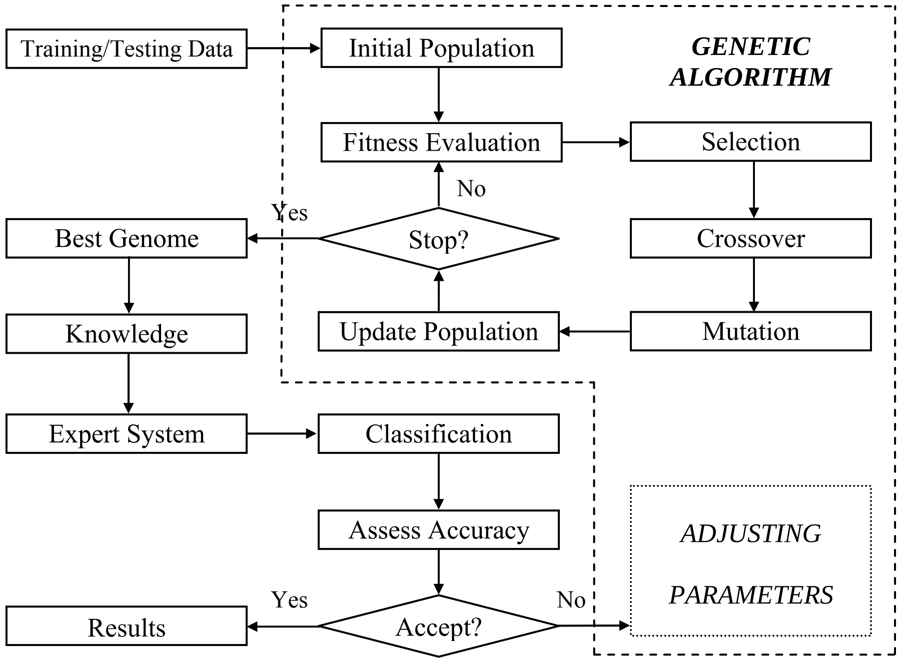


Figure 4.1. The workflow for optimizing genetic parameters configuration in genetic algorithm-based land-cover and land-use classification

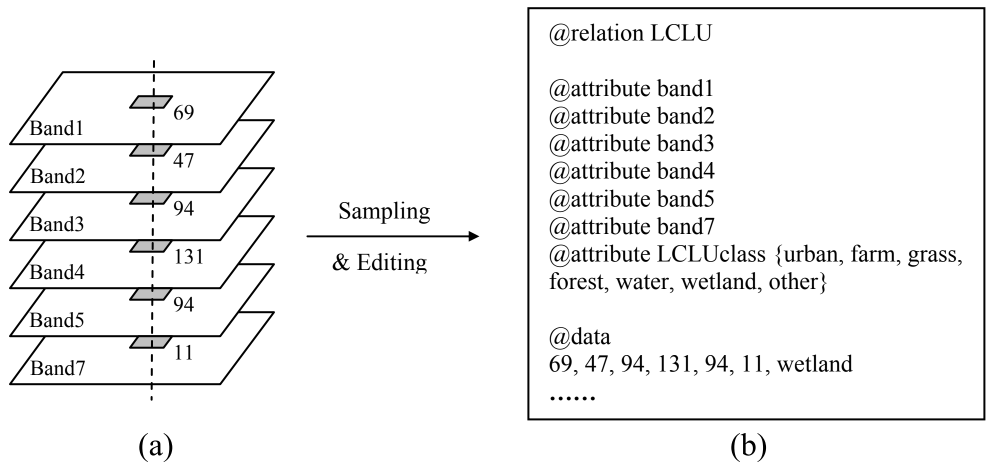


Figure 4.2. (a) Randomly sampled training data and (b) Edited data file for optimizing genetic parameters configuration

Table 4.1. Translating a portion of decision tree into classification rules
(during optimizing genetic parameters configuration)

<table>
<tr><th>Decision Tree</th><th>Classification Rules</th></tr>
<tr><td>if band4 <= 76 then
| if band5 <= 94 then
| | if band3 <= 94 then
| | | if band1 <= 69 then
| | | | if band2 <= 47 then
| | | | | if band 6 <= 11 then
| | | | | | if band4 <= 131 then
| | | | | | | - wetland
| | | | | | + - forest
| | | | | + - grass
| | | | + - urban
| | | + - urban
| | + - other
| + - farm
|……</td><td>Rule 1. if band4 <= 76, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, band6 <= 11, and band4 <= 131, then class = wetland;
Rule 2. if band4 <= 76, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, band6 <= 11, and band4 > 131, then class = forest;
Rule 3. if band4 <= 76, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, and band6 > 11, then class = grass;
Rule 4. if band4 <= 76, band5 <= 94, band3 <= 94, band1 <= 69, and band2 > 47, then class = urban;
Rule 5. if band4 <= 76, band5 <= 94, band3 <= 94, and band1 > 69, then class = urban;
Rule 6. if band4 <= 76, band5 <= 94, and band3 > 94, then class = other;
Rule 7. if band4 <= 76, and band5 > 94, then class = farm.
……</td></tr>
</table>

If only one year's DOQQs, for example 1998 DOQQs, were used for classification accuracy assessment, it may lead to error because of the temporal differences between the acquisition date of 2000 Landsat Thematic and 1998 DOQQs. In order to prevent this error, both 1998 DOQQs and 2004 DOQQs were used simultaneously. The assumption is that any pixel that does not change on both DOQQs will remain the same on the Landsat TM image also. Such pixels will be used for accuracy assessment. During the procedure of accuracy assessment, if the comparison of the same pixel on two DOQQs suggests a LCLU change, the pixel will be abandoned and new points will be selected until 1000 pixels are defined.

4.2. Results

This study only focuses on a subclass of GAs characterized by the following five genetic factors. The default setting up for carrying out this experiment includes: train / testing data size (10000 / 5000), spatial resolution (30 meters), generation size (2000), population size (1000), replacement rate (50%), crossover rate (99%), and mutation rate (5%). When carrying out an experiment on a parameter, the rest of the parameters remain at default values (Table 4.2).

A population will converge when most individuals in the population are identical, or in other words, the diversity of the population is minimized (Louis and Rawlins, 1993). In general, different populations converge to different levels of fitness values. A population with high diversity will eventually converge to a global optimum fitness. A population with low diversity will quickly converge to a local optimum fitness. Models based on global optimum can capture most data, but models based on local optimum cannot capture most data. The experiment finds the relationship among the different levels of convergence, genetic / environmental parameters, and classification accuracy. The genetic parameters include generation size, population size, generation gap, crossover rate, and mutation rate. The levels of these parameters that are too low, too high, or at middle will affect the performance of genetic algorithms differently.

Table 4.2. The measurement of the performance of genetic algorithms

	Case Id	Size or Rate	Best Genome Fitness	Average Genome Fitness	Best Genome Size	Training Data Classification Accuracy	Testing Data Classification Accuracy
Generat. Size	**G1**	**100**	**0.5463**	**0.5444**	**75**	**0.7393**	**0.7330**
	G2	500	0.6134	0.6129	167	0.7843	0.7764
	G3	1000	0.6315	0.6309	171	0.7958	0.7875
	G4	2000	0.6412	0.6408	135	0.8015	0.7878
	G5	**5000**	**0.6487**	**0.6485**	**135**	**0.8062**	**0.7882**
Popul. Size	**P1**	**250**	**0.6011**	**0.6010**	**109**	**0.7758**	**0.7686**
	P2	500	0.6315	0.6309	159	0.7957	0.7858
	P3	750	0.6306	0.6301	173	0.7953	0.7833
	P4	1000	0.6412	0.6408	135	0.8015	0.7878
Crossover Rate	**C1**	**19%**	**0.5820**	**0.5820**	**73**	**0.7631**	**0.7581**
	C2	39%	0.5938	0.5938	63	0.7708	0.7635
	C3	59%	0.6160	0.6160	95	0.7852	0.7762
	C4	79%	0.6173	0.6173	105	0.7861	0.7786
	C5	99%	0.6412	0.6408	135	0.8015	0.7878
Mutation Rate	M1	0.01%	0.6329	0.6326	149	0.7965	0.7814
	M2	0.1%	0.6333	0.6326	173	0.7970	0.7867
	M3	0.5%	0.6412	0.6408	135	0.8015	0.7878
	M4	1%	0.6288	0.6283	141	7938	0.7874
	M5	10%	0.6008	0.6000	99	0.7755	0.7685
	M6	**100%**	**0.3663**	**0.3663**	**39**	**0.6053**	**0.6040**
Generat. Gap	**R1**	**10%**	**0.5990**	**0.5976**	**121**	**0.7745**	**0.7745**
	R2	25%	0.6212	0.6202	141	0.7890	0.7781
	R3	50%	0.6412	0.6408	135	0.8015	0.7878
	R4	75%	0.6400	0.6398	155	0.8010	0.7873

4.2.1. Number of Generations

Generation is a procedure to produce a new population (Goldberg, 1989), and a generation is one evolution cycle (Freeman, 2002). When the evolution cycle is constant, sufficient time is a prerequisite for a population to achieve global optimum. As expected, the larger the number of generation, the more computational resources (*e.g.*, CPU time and RAM space) are needed.

Both Figures 4.3 and 4.4 demonstrate the convergence process of a population during the evolution of 5000 generations. The best genome fitness and average population fitness increase while the generation number increases, and a diminishing returns effect can be observed. During the early evolutionary stage (1-500 generations), the various convergence rates were observed, and the population quickly approaches a relatively fitness level. After the first 1000 generations, the low and similar convergence rates within a comparable scale from 1000 to 2000 generations and different final fitness value under different generation sizes were achieved, and the population showed a slow evolutionary progress.

This result suggests that the convergence rate change during entire evolutionary process and implies that sufficient time is a prerequisite for evolution. It also indicates that a population could not converge to a global optimum when the amount of generation is too small. For example, when the amount of generation was 100, the best genome fitness and average genome fitness was only 0.5463 and 0.5444, respectively (Table 4.2). The classification accuracy of training data and testing data was only 73.93% and 73.30%, respectively (Table 4.2). Unlike the population with a small generation amount, a population with a large generation number has enough time for a successful evolution. A large numbers of generations usually leads to a population with high fitness and global optimum. However, it is going to cost large amount of computational resources.

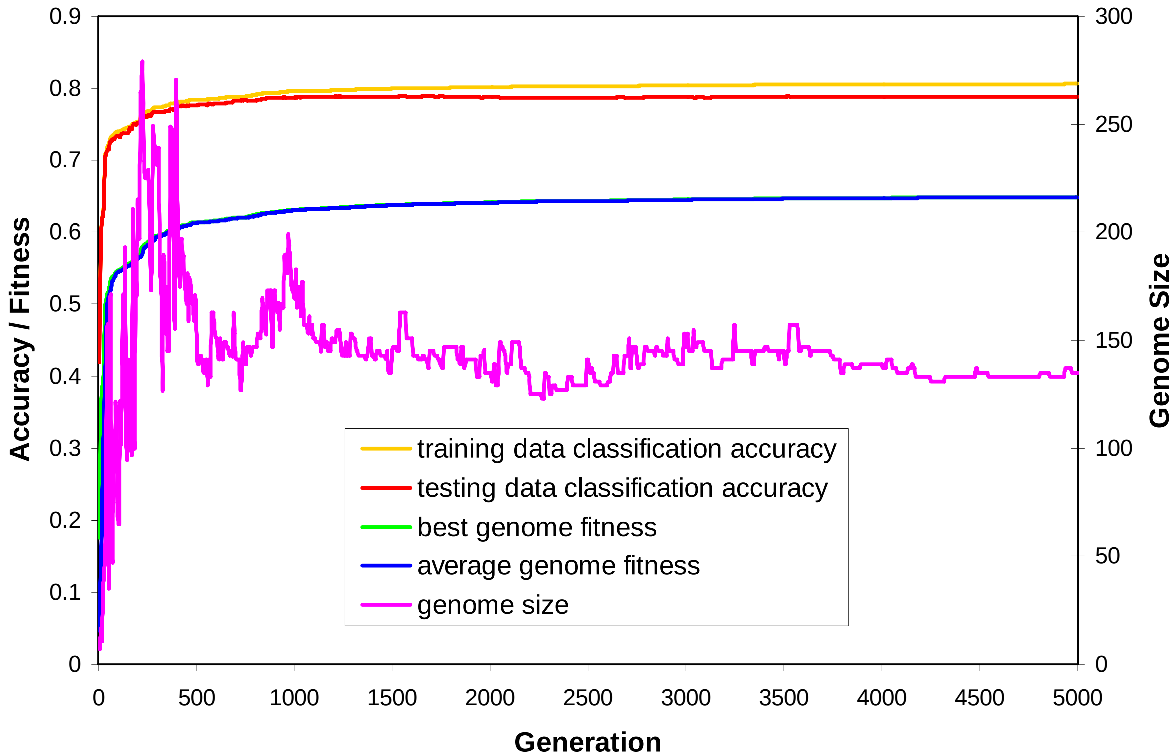


Figure 4.3. Relationship among accuracy / fitness, genome size and number of generation

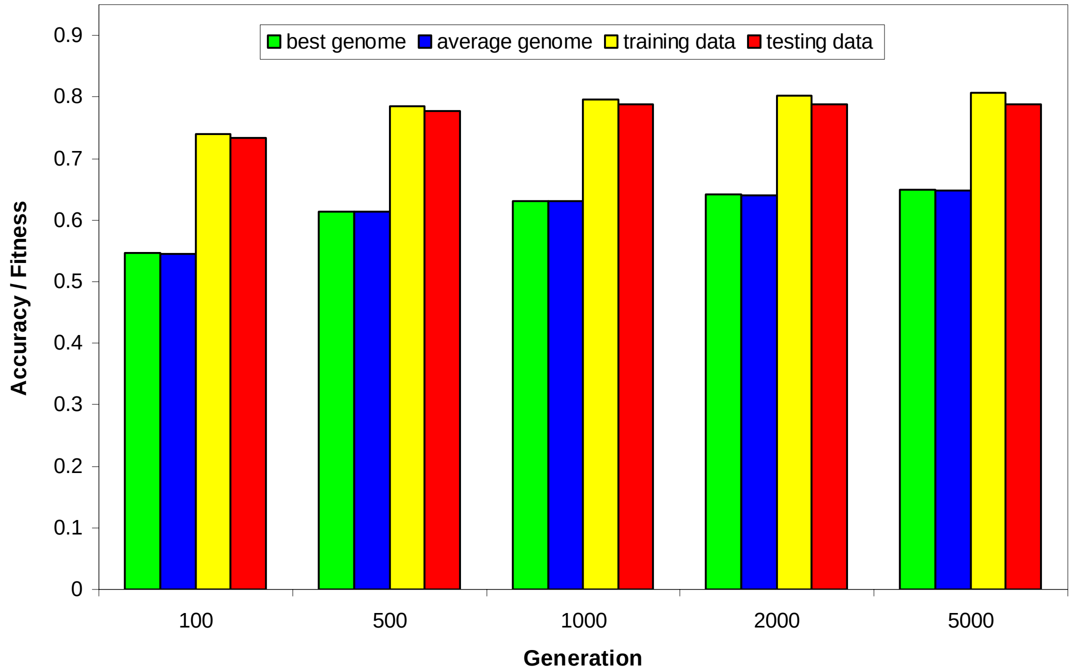


Figure 4.4. Comparing the accuracy / fitness of population
when the number of generations changes

4.2.2. Initial Population Size

After inputting training / testing data derived from the Landsat image into GATree 2.0, a set of solutions was generated. A solution represents an individual, and a number of individuals (or solutions) consists of a population. Each solution is a classification model that can be used for classifying the Landsat image. The population is a group of individual solutions (or decision trees) with the same or similar characteristics. Although they are within same population, different individuals have different genomes and fitness. Population size is one of most important factors impacting the performance of GAs. It affects both the ultimate performance and the efficiency of GAs, and GAs usually perform poorly with small population (Grefenstette, 1986). Population size differs from training / testing data size. If the population size is too small, the population will not have a large gene pool for successful evolution. It means this population provides only insufficient samples for most hyperplanes. GAs can perform a more efficient search if population is large because this population has more chance to cover various individuals from a large number of hyperplanes (Grefenstette, 1986).

In this research, relationships among population size, fitness, classification accuracy, and generation size are given in Figures 4.5 and 4.6. Generally, a large population size provides high diversity so that the population has more chance to develop global optimized gene pool during the evolution. Figure 4.5 shows that a smaller population converges to a lower level global optimum (*e.g.*, best/average fitness 0.6011/0.6010), while a larger population converges to a higher level global optimum (*e.g.*, best/average fitness 0.6412/0.6408). Also, a larger population has higher classification accuracy than a smaller population (Figure 4.6). However, a large population size may not always yield a greater benefits. In this study, the best /average genome fitness increases while population size increases from 250 to 500. But after the size of population becomes larger than 500, the best / average genome fitness only increases slightly (Table 4.2).

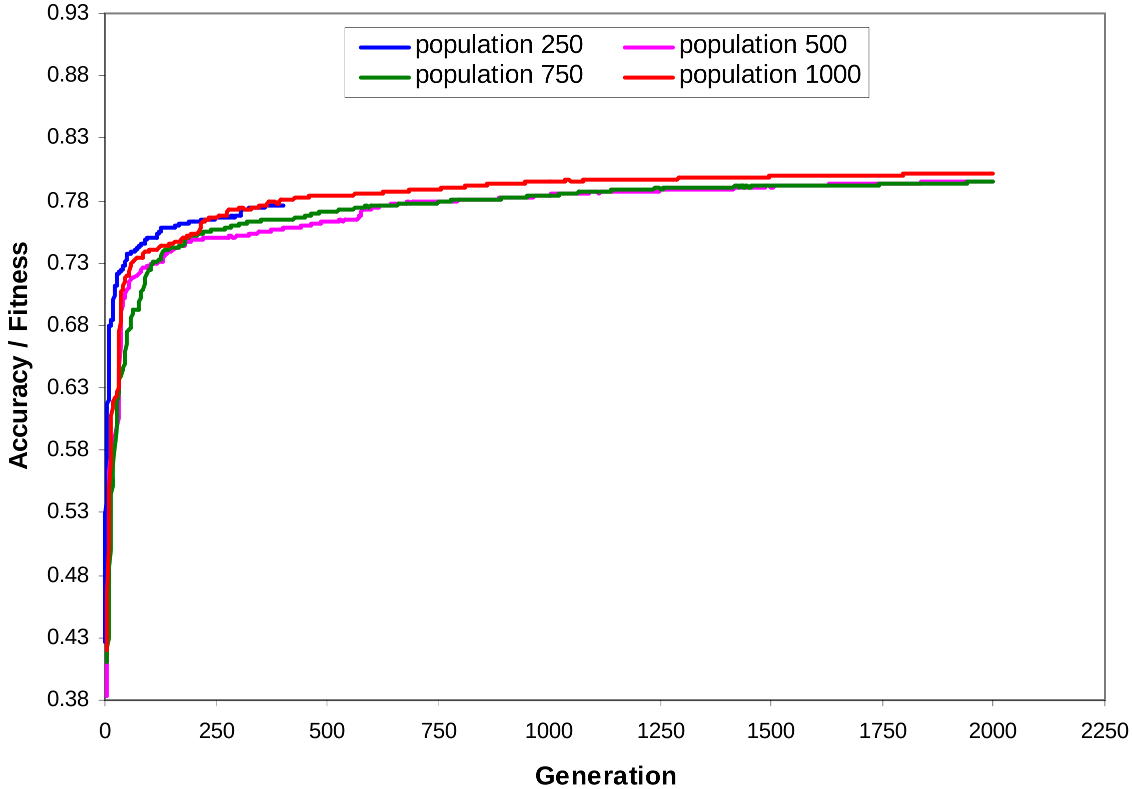


Figure 4.5. Relationship among accuracy / fitness, population size, and number of generation

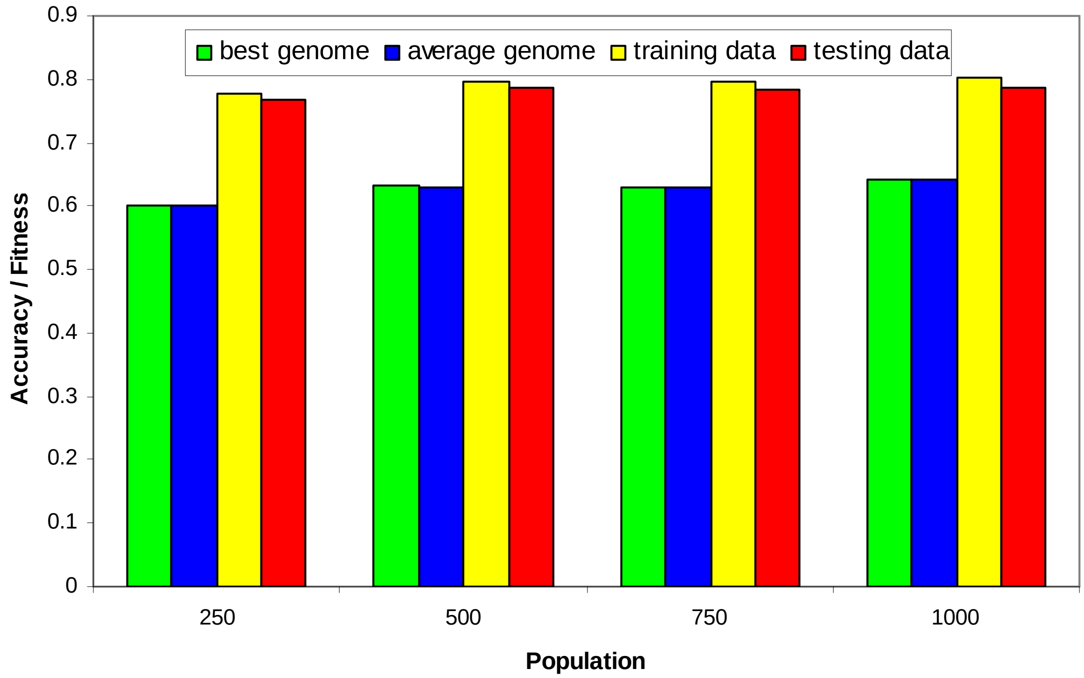


Figure 4.6. Comparing the accuracy / fitness of population
when population size changes

The possible reason for such phenomena is resource limitation (*e.g*., small RAM) or population crowding. Although a large population can prevent premature convergence to local solutions, it requires more time to finish per generation so evolution will be slow.

4.2.3. Crossover Rate

The crossover is the procedure to generate children from two parents. Crossover is also an important genetic operator because it introduces biodiversity into the population. The significance of crossover rate in controlling the performance of GAs has been recognized theoretically and empirically (Grefensette, 1986; Srinivas and Patnaik, 1994). The crossover rate controls the frequency with which the crossover operator is applied. Theoretically, when both crossover rate (C) and population size (N) are given, for each new population, there are $C*N$ selected individuals for crossover. The higher the crossover rate, the quicker the new individuals are introduced into the population (Grefensette, 1986; Srinivas and Patnaik, 1994).

The results in both Figures 4.7 and 4.8 illustrate that higher crossover rate leads to faster/higher level convergence and vice versa. When crossover rate is 0.19 and 0.99, the best genome fitness is 0.5820 and 0.6412, respectively (Table 4.2). The reason for these results is that a smaller crossover rate cannot introduce enough new individuals into the population rapidly. Therefore, it cannot lead to the high level global optima within a given time.

Some authors argue that if the crossover rate is too high, individuals with high fitness are discarded faster than selection can produce (Grefensette, 1986; Srinivas and Patnaik, 1994).
In such situations, the population prematurely converges to the local optima, because a higher crossover rate tends to disrupt the individuals selected for reproduction at a high rate. This characteristic is important in a small population because high-fitness individuals are more likely to quickly dominate a small population (Grefensette, 1986). It may be different if the population is large.

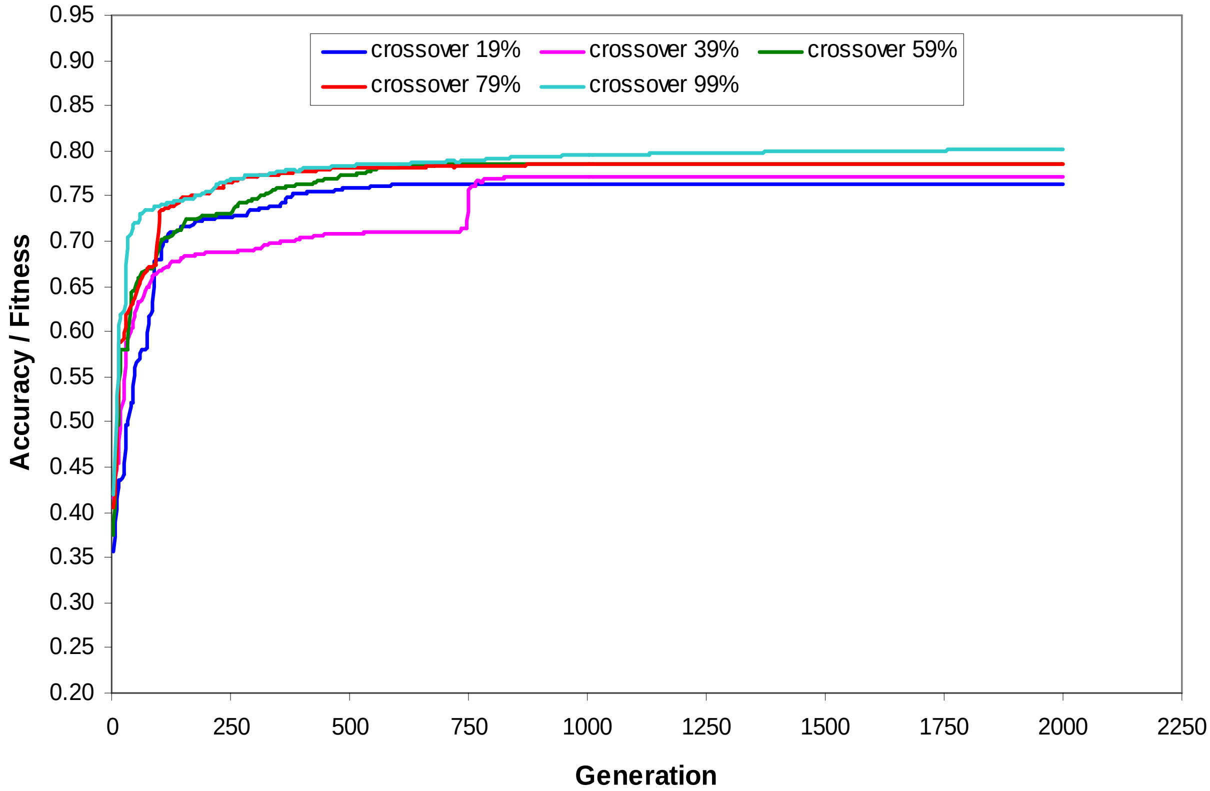


Figure 4.7. Relationship among accuracy / fitness, crossover rate, and number of generation

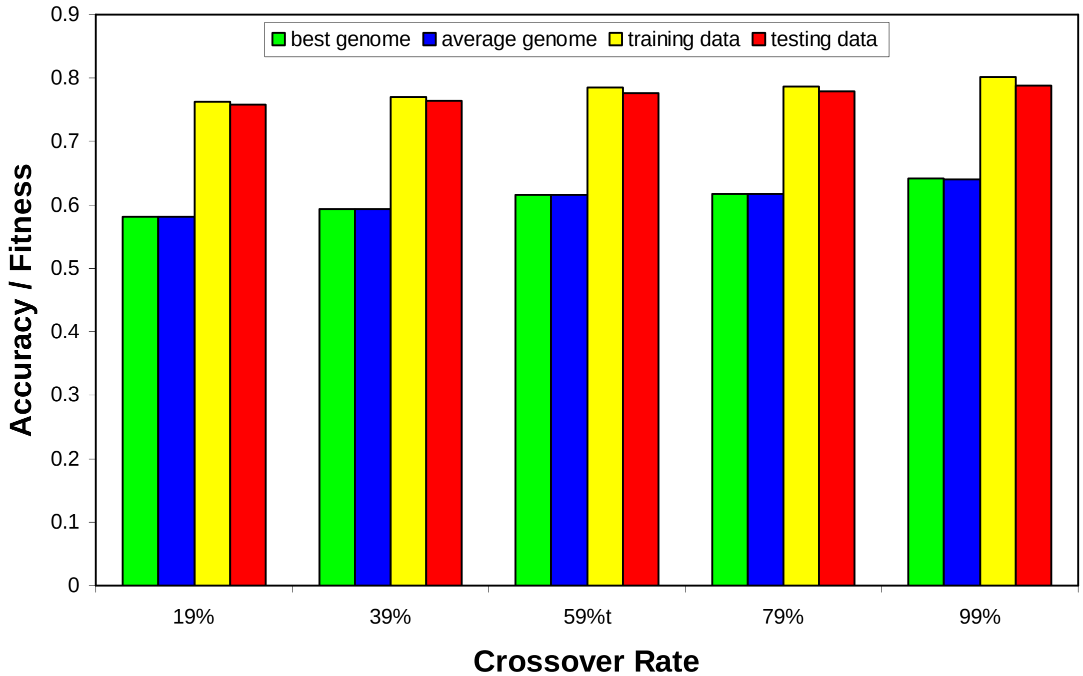


Figure 4.8. Comparing the accuracy / fitness of population when the rate of crossover changes

The results from this study indicate that the highest crossover rate 0.99 does not lead to premature convergence in a population with 1000 individuals (Table 4.7, Figure 4.8). When Grefensette (1986) suggested that crossover plays an important role in preventing premature convergence in smaller populations, he dealt only with population sizes from 10 to 160 individuals. So, it can be argued that a very high crossover rate may not necessarily cause premature convergence in a large population.

4.2.4. Mutation Rate

Mutation is a procedure to change a selected individual genome thereby introducing biodiversity into the population. The role of mutation in GAs is to restore lost or unexplored genomes into the population and to prevent premature convergence or local optima. Similar to the study of crossover rate, the significance of mutation rate in controlling the performance of GAs has been recognized both theoretically and empirically (Grefensette, 1986; Srinivas and Patnaik, 1994). The mutation rate controls the frequency with which the mutation operator is applied. Figures 4.9 and 4.10 show the relationship between best genome fitness and generation size. When the mutation rate is large, such as 100%, the population quickly converges to a lower level optimum (both the final best and average fitness are 0.3663). A high-rate mutation destroys genomes that have either high or low fitness, and introduces many new but not necessary high fitness genomes into the population. Although the higher rate of mutation usually leads to an essentially random search (Grefensette, 1986), it usually drives a population into an unstable situation. The higher-rate mutation usually leads an essentially random search (Grefensette, 1986). Of course, the result of such a search will not be desirable.

With a smaller mutation rate, such as 0.01%, a population slowly converges to a higher-level optimum (the final best / average fitness is 0.6329 and 0.6326 respectively). This is because a low mutation rate prevents any position on a genome from being changed and introduces few

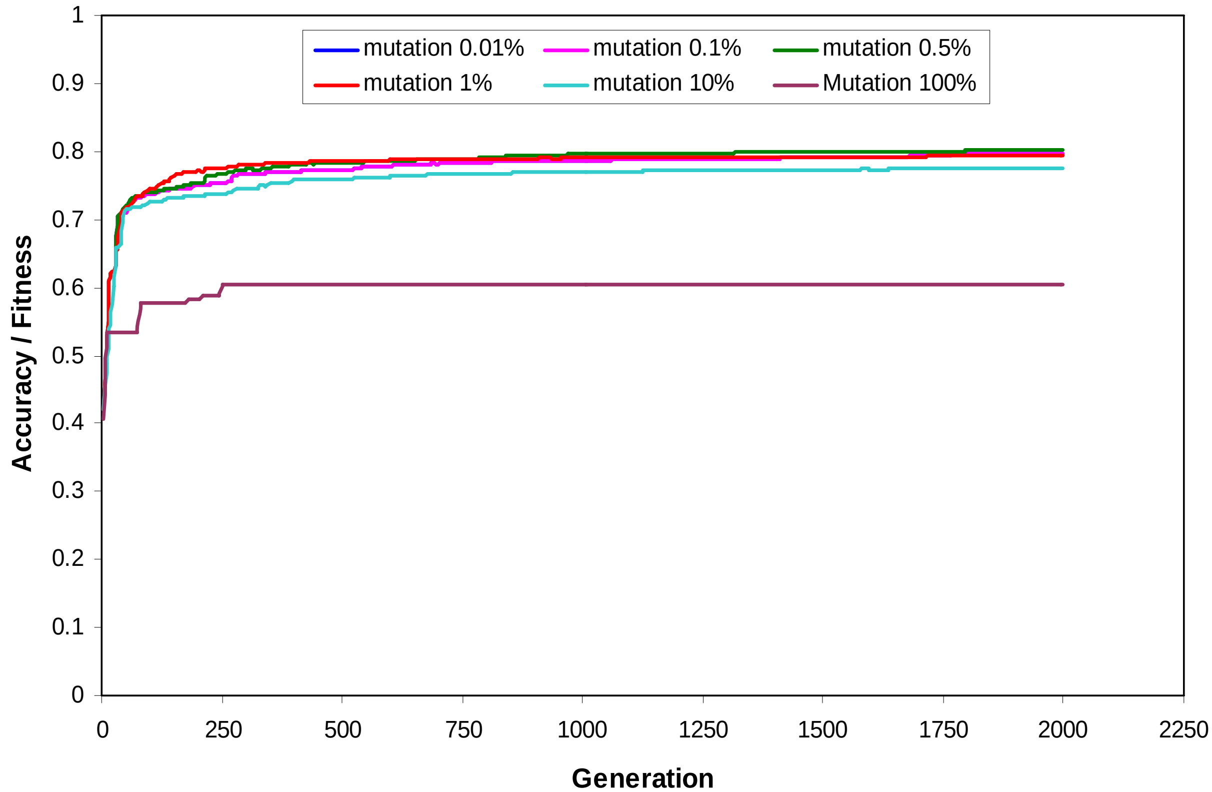


Figure 4.9. Relationship between accuracy / fitness, mutation rate, and number of generation

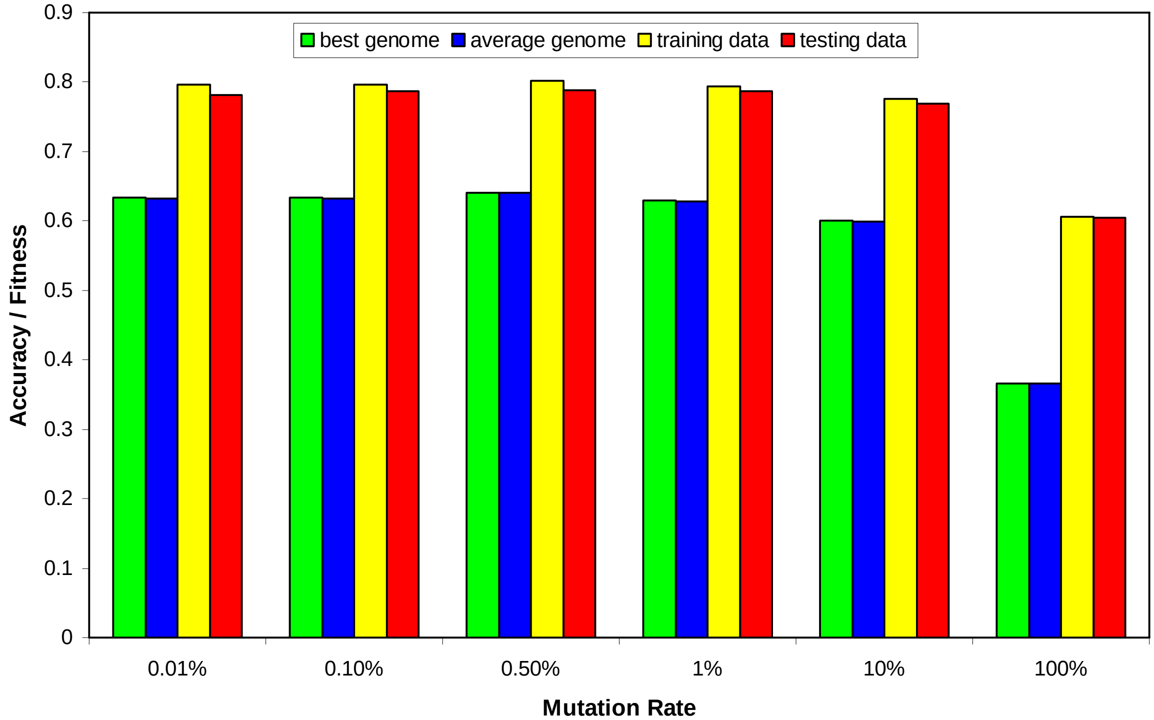


Figure 4.10. Comparing the accuracy / fitness of population when mutation rate changes

new genomes into the population (in this case, only one new genome or decision tree introduced per generation). Although high-fitness genomes are more likely to be saved in this case, the biodiversity of the population will be affected. Also, populations with a low mutation rate usually require more time to converge to the global optimum.

4.2.5. Generation Gap

The concept of generation gap, or so-called replacement rate, was introduced into GAs-based study by De Jong and Sarma (1993). They empirically evaluated the performance of GAs with overlapping populations and found that a population loses high-fitness genomes when the generation gap is small. Small generation gap can cause poor search performance (De Jong and Sarma, 1993). Cheng *et al*. (1996) believe that a very small generation gap may lead to premature convergence.

Grefenstette (1986) stated that the generation gap controls the percentage of the population to be replaced during each generation. A value of 100% means that the entire population is replaced during each generation, and a value of 50% means that half of the individuals in each population survive into the next generation (Grefenstette, 1986). Grefenstette (1986) also argued that the large generation gap generally improves performance of GAs. However, the study done by Wang and Cao (2002) suggested that a very large generation gap may decrease population diversity and limit search space.

Figure 4.11illustrates that a population with a higher generation gap (*e.g.* 75%) leads to a higher convergence rate during the early stage of evolution (about first 500 generations), then the evolutionary process of such a population is slowed down and surpassed by populations with a lower generation gap (*e.g.* 50%). This observation does not support Grefenstette's (1986) argument. Grefenstette's experiment was done based on a maximum population size 160 and a maximum generation number of 20, so the behavior of genetic algorithms after 20 generations

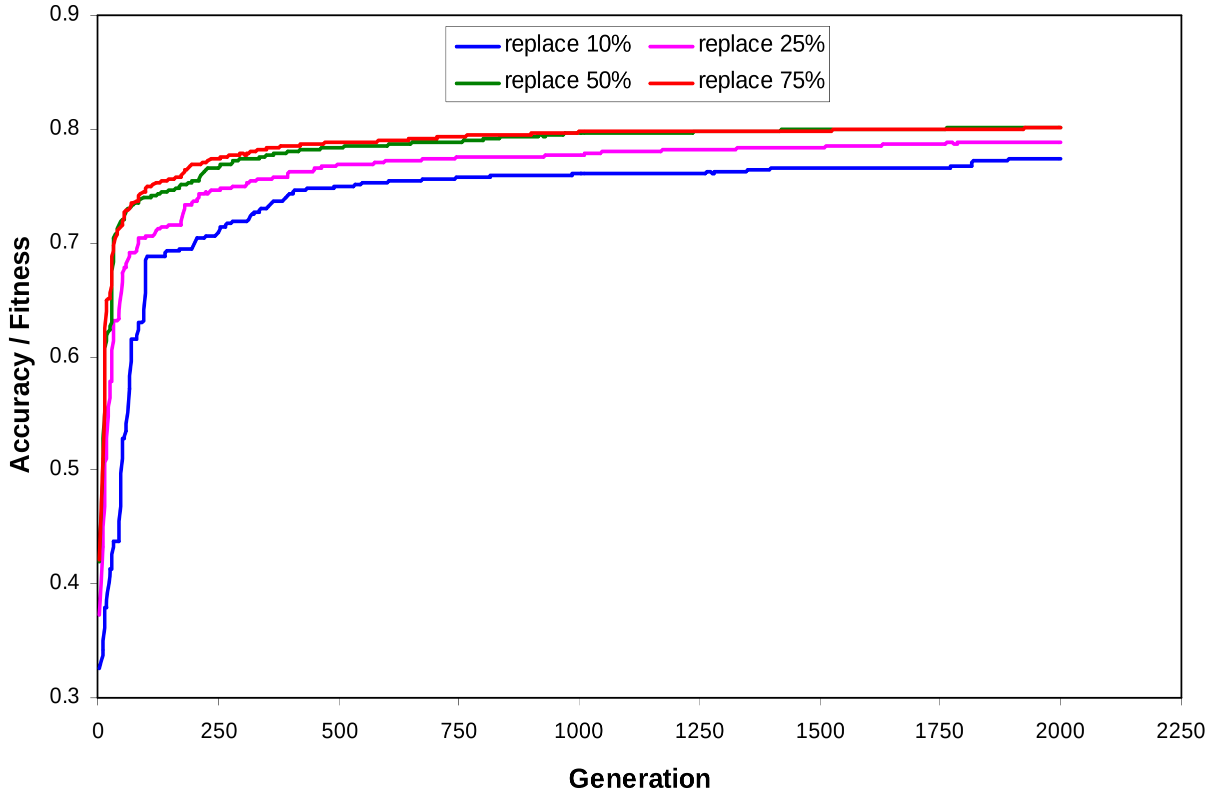


Figure 4.11. Relationship among accuracy / fitness, generation gap, and number of generation

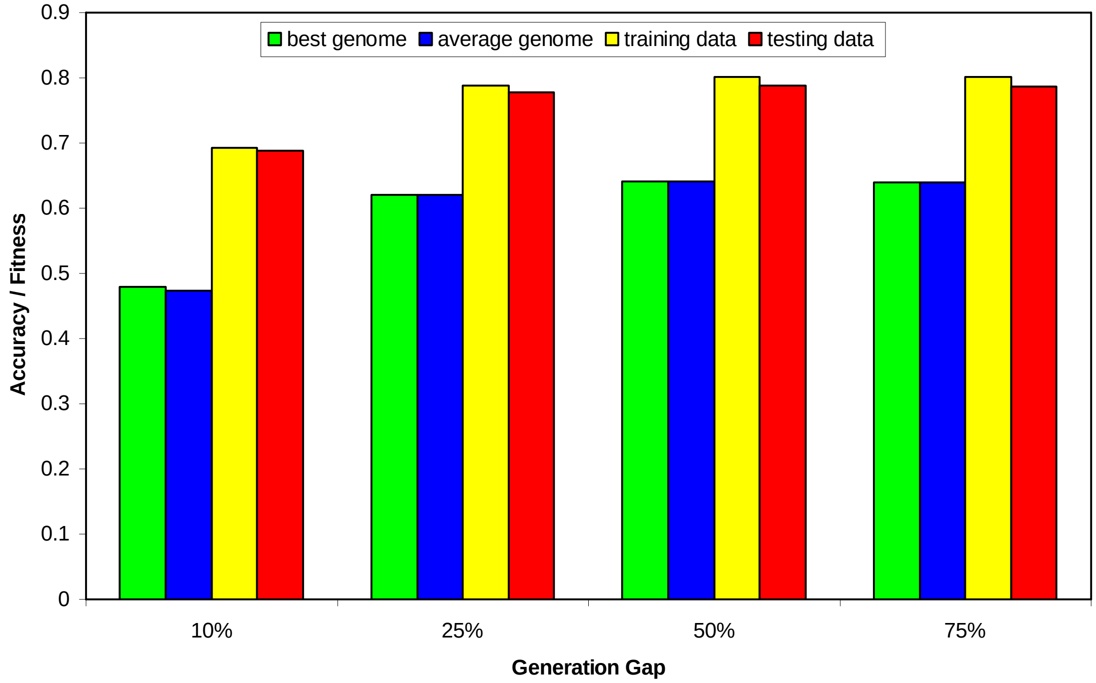


Figure 4.12. Comparing the accuracy / fitness of population when generation gap changes

was not observed. Figure 4.12 and Table 4.2 compare fitness and classification accuracy among populations with different generation gaps. During a 2000-generation evolution, a population with a 50% generation gap has the highest best/average genome fitness and training/testing data classification accuracy. A population with 10% generation gap has lowest best/average genome fitness and training/testing data classification accuracy, and the difference between populations with 25% and 75% generation gap is small. These results indicate that the performance of GAs will be poor when the generation gap is too small or too large. If a population has a lower fitness and smaller generation gap, it has less chance to swap high fitness genomes into the next generation. The performance of GAs can be improved by a large generation gap during early stage of evolution (*e.g.* increase convergence speed), but it will not necessarily be improved during the entire evolution when the generation gap is large. Because with a large generation gap, for example 75%, most original individuals in the population will be replaced by new individuals. This will cause the population to be unstable and eventually leads to the negative effects of genetic drift (*e.g.*, allele loss or gene loss) and a local optimum. Another drawback of large generation gap is that it requires much time (*e.g.* 6293 minutes per experiment).

4.2.6. Land-Cover and Land-Use Classification

The experiment designed in this chapter is used to determine the optimal configuration of genetic parameters. When the fitness of population is low, the classification accuracy of training data and testing data will be low. For example, both the best and average genome fitness of a population in case M6 are only 0.3663, and its training / testing data classification accuracy is low as 60.53% / 60.40% (Table 4.2). When the fitness of a population is high, the classification accuracy of training data and testing data will be high. For example, both the best and average genome fitness of a population in case G5 are 0.6487, and its training / testing data classification accuracy is also low at 80.62% / 78.82% (Table 4.2).

Usually, both the high fitness of genome and the high classification accuracy of training / testing data indicate a better performance of the GA. Based on the performance of GAs, the following cases were selected for carrying out a real image data classification: 1) Case G1 has a very small generation amount (100); 2) Case P1 has a very small population size (250); 3) Case C1 has a very low crossover rate (19%); 4) Case M6 has a very high mutation rate (100%); 5) Case R1 has a very small generation gap (10%); and 6) Case G5 has an optimal configuration (generation amount 5000, population size 1000, crossover rate 99%, mutation rate 0.5%, generation gap 50%).

Both Table 4.3 and Figure 4.13 show the relationship between genetic algorithms performance and LCLU classification accuracy. When the classification rules were developed from an effective genetic algorithm, the classification accuracy of the image data was high. Otherwise, the classification accuracy of the image data was low. The comparison of Case M6 and Case G5 provides a good example. When the population fitness and training / testing data classification accuracy is high, the classification accuracy of real data is also high (*e.g.*, Case G5, 84.70%), and likewise, when the population fitness and training / testing data classification accuracy is low, the accuracy of real data LCLU classification is low (*e.g.*, 22.70% in Case M6) (Table 4.3).

4.3. Discussion

The impact of genetic algorithms performance on LCLU classification includes at least three aspects: 1) the impact of premature convergence / local optimization; 2) the impact of a very low convergence rate and an unstable population; and 3) the impact of a small number of generations and a small population size. All of these are closely related to the configuration of genetic parameters. In this section, these four issues were discussed, and a recommendation of GA parameters configuration was also provided.

Table 4.3. Classification accuracy assessment for genetic parameter configurations*

Designed Experiments		**Generation Amount**	**Population Size**	**Crossover Rate**	**Mutation Rate**	**Generation Gap**	**Optimum Setting**
Case ID**		G1	P1	C1	M6	R1	G5
Urban	Prod. Accu.	53.41%	86.36%	56.82%	17.61%	68.18%	80.68%
	User Accu.	53.41%	63.07%	75.76%	63.27%	66.67%	88.20%
Farm	Prod. Accu.	92.46%	92.96%	94.97%	93.47%	90.45%	92.46%
	User Accu.	60.73%	76.45%	66.08%	45.81%	76.92%	78.30%
Grass	Prod. Accu.	3.26%	--	28.26%	--	28.26%	38.04%
	User Accu.	75.00%	--	57.78%	--	72.22%	54.69%
Forest	Prod. Accu.	96.68%	93.36%	92.89%	6.52%	94.31%	92.89%
	User Accu.	82.59%	89.55%	89.50%	2.31%	88.44%	85.96%
Water	Prod. Accu.	93.94%	96.97%	100.00%	--	96.97%	87.88%
	User Accu.	100.00%	100.00%	97.06%	--	100.00%	100.00%
Wetland	Prod. Accu.	85.53%	89.36%	92.77%	6.06%	92.34%	90.64%
	User Accu.	97.10%	97.67%	95.61%	1.40%	95.59%	94.25%
Other	Prod. Accu.	4.76%	85.71%	80.95%	--	85.71%	90.48%
	User Accu.	100.00%	100.00%	77.27%	--	52.94%	67.86%
Overall Accuracy		74.90% /	82.60%	81.20%	22.70%	82.40%	84.70%
Overall Kappa		0.6876	0.7846	0.7682	0.1093	0.7834	0.8119

* 1000 reference points
** The configurations of genetic parameter in different case are listed in Table 4.2

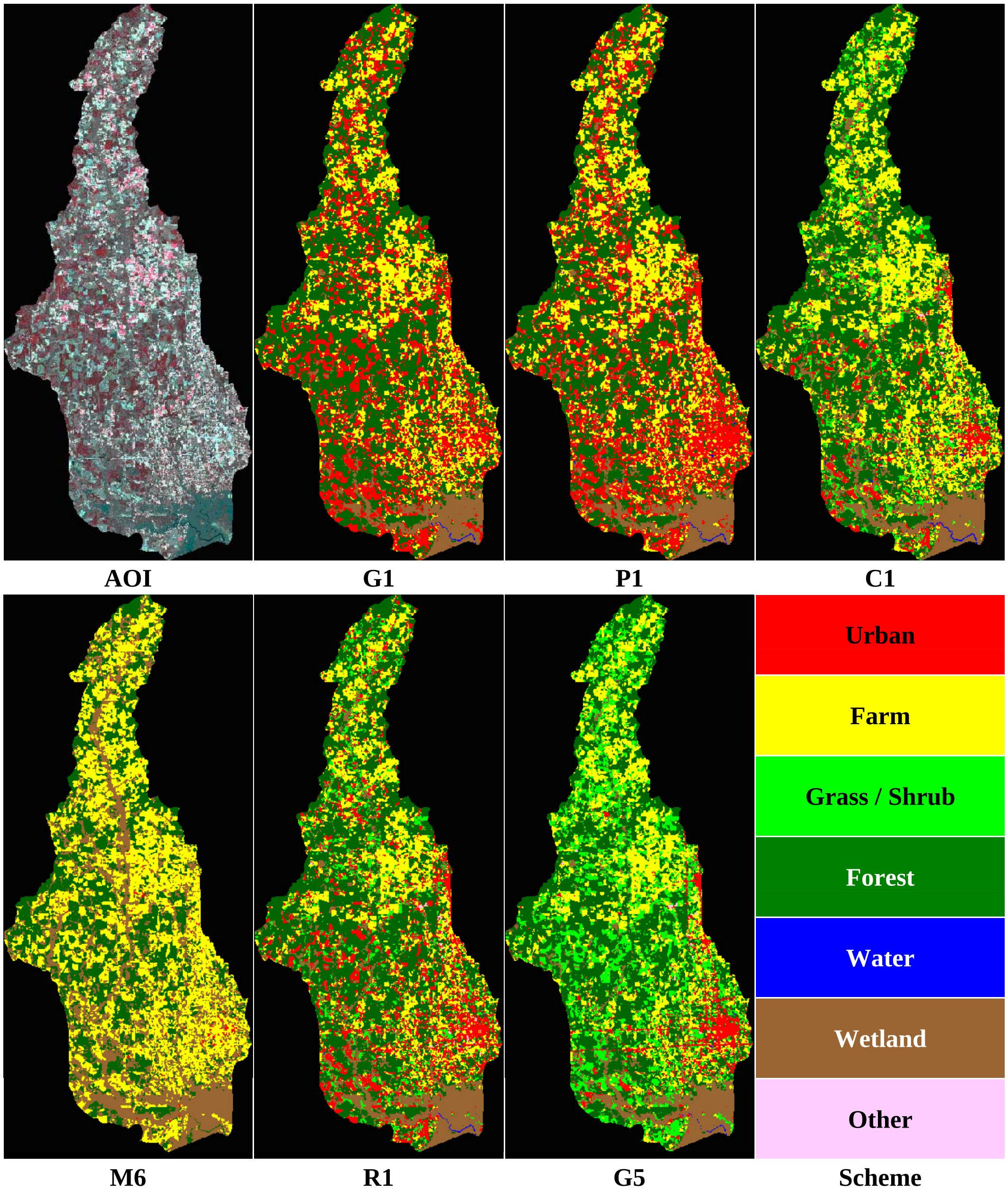


Figure 4.13. Genetic algorithms-based land-cover and land-use classification using different genetic parameter configurations listed in Table 4.2

4.3.1. Impact of Premature Convergence and Local Optimization

Convergence is assumed to be achieved when there is no increase in the maximum fitness of population with an increased number of generations (Chambers, 2000). It can also be regarded as the state when most of the population is identical and diversity is minimized (Louis and Rawline, 1993). There are three types of convergence: 1) rapid convergence, 2) moderate rate of convergence, and 3) slow convergence (Chambers, 2000). Each of these has different rates of convergence, and the best individual requires a different amount of time to reach best fitness.

Premature convergence / local optimization is a phenomenon that cannot be ignored in GAs. It has two aspects: 1) all individuals in a population become the same so that the evolution stops; and 2) the population cannot converge to any level of fitness because the better individuals are always eliminated (Cheng *et al*., 1996). The major reason may be that individuals with very high fitness are assigned during the early evolution stage. Fitness-based natural selection eliminates other individuals so that most individuals are the same. Crossover occurs between the same type of individuals cannot produce new offspring. Mutation may generate some high-fitness individuals, but they have high probability to be abandoned because of the small number of individuals. Eventually, most individuals become the same (Cheng *et al*., 1996). Although it has high fitness, this population (a set of solutions) cannot be used to solve real problems because it is a local optima.

One of the cases shown in Table 4.2 is about premature convergence / local optimization. Case M6 has the highest mutation rate – 100%. This population becomes mature around 250 generations and converges to a low fitness level of 0.3663 (Table 4.2). Its training / testing data classification accuracy is just 0.6053 / 0.6040. Its real data classification accuracy is only 22.70%. This is a typical example of premature convergence. When the rate of mutation is very high, the population (models) cannot be stable, and its good characteristics cannot be efficiently transmitted from one generation to the next. When using the model (best individual) to classify testing data, the accuracy was not high (60.40%), and when using the model to classify a satellite image the accuracy was even worse (22.70%) (Table 4.3). This result suggests that premature convergence / local optimization can impact the accuracy of LCLU classification negatively. In order to prevent premature convergence / local optimization, the genetic parameters must be set up properly.

4.3.2. Impact of Low Convergence Rate and Unstable Population

Low convergence can be caused by introducing very few new individuals into population in each generation. The possible reason could be a low crossover rate, low mutation rate, and small generation gap. Table 4.2 shows that a low crossover rate leads to low fitness and low classification accuracy, as in Cases C1 and C2. It also shows that a small generation gap leads to low fitness and low classification accuracy, as in Case R1. Although a low mutation rate cannot introduce enough new individuals into the population, if given enough time, it will lead to global optimum (Grefensette, 1986). This may be the reason why Case M1 has high fitness and classification accuracy.

A higher mutation rate can introduce many new individuals to the population and prevent a population from premature convergence. But if it is too high (*e.g.* 0.1 and 1.0), the random search will dominate the behavior of GAs, and the optimization capability will be lost. Therefore, a very high mutation rate could lead to an unstable population that usually converges to a low fitness level. Similar to the situation with a high mutation rate, when the generation gap is large (*e.g.*, 25% and 50%), it introduces new individuals into the next population and prevents population from the problem of premature convergence. But if generation gap is too large (*e.g.,* 75%), the random search will dominate the behavior of GAs, and the optimization capability will be lost. Therefore, a very large generation gap could lead to unstable population that usually converges to a low fitness level. The results in Table 4.3 indicate that either premature convergence or an unstable population will lead to low classification accuracy on either testing data or the entire study area, as in Case M6.

4.3.3. Impact of Small Number of Generations and Small Size of Population

A small number of generations, such as 100, cannot provide enough time for evolution. A population with small number of generations usually does not develop well. It has poor fitness and classification accuracy on both testing data and satellite imagery. Cases G1 and G2 belong to this scenario. When the generation size is 100, the best fitness is only 0.5463, testing data classification accuracy is only 0.7259, and satellite image classification accuracy is only 0.7330.

A small population size, such as 250, cannot provide enough diversity for better evolution. A population with a small diversity usually does not develop well. It has poor fitness and classification accuracy on both testing data and satellite imagery. Case P1 belongs to this scenario. When the population size is 250, the best fitness is only 0.6011, testing data classification accuracy is only 76.86%, and satellite image classification accuracy is only 82.60%. These results suggest that small generation size and small population size can decrease the capability of GAs. The performance of GAs can be improved by increasing generation amount and population size.

Both Table 4.3 and Figure 4.13 show the relationship between GA performance and LCLU classification accuracy. Generally, when the fitness of a population is high, the classification accuracy of training data and testing data will be high. The high fitness of best / average genome and the high classification accuracy of training / testing data indicate the good performance of GAs. However, if premature convergence or local optimization occurs, the situation will be complex, any model based on premature convergence or local optimization will not perform well in terms of LCLU classification.

4.3.4. Recommendation of a Genetic Parameters Configuration

Based on the above results and discussion, the following recommendations can be made. To achieve better classification results, the genetic parameters range should be: generation 2000 – 5000, population 1000 – 2000, crossover rate 69% – 99%, mutation rate 0.1% – 0.5%, and generation gap 25% – 50%. The best scenario from this study is: generation 5000, population 1000, crossover rate 99%, mutation rate 0.5%,, and generation gap 50%.

Table 4.3 and Figure 4.13 show the results of satellite image classification accuracy for different cases. The recommended case has the highest classification accuracy for both testing data and satellite image (78.82% / 84.70%). These results suggest that the performance of GAs and LCLU classification can be improved by adjusting genetic parameters.

4.4. Hypothesis # 1 Review

The first research question as presented in Chapter 1 was How do genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generation gap) impact the accuracy of GA-based LCLU classification. The first research hypothesis presented in Chapter 1 was stated as follows:

Ha: Different numbers of generations, population size, crossover rate, mutation rate, and generation gap can increase / decrease the accuracy of GA-based LCLU classification, respectively.

The impacts of various genetic parameters on the performance of GA-based LCLU classification, as the above experimental results show, are complex and significant. For example, when generation gap is smaller than 50%, the performance of GA-based LCLU classification can be improved as the generation gap increases. However, it becomes bad as the generation gap increases from 50% to 75%. The results of the experiment indicate that the first hypothesis can be accepted. It means that different number of generation, population size, crossover rate, mutation rate, or generation gap can impact GA-based LCLU classification, respectively. In general, the performance can be improved by increasing the number of generations, the size of population, and the rate of crossover. It can be improved by decreasing the rate of mutation also.

4.5. Summary

As it has been stated that the primary goal of this study is to investigate premature convergence / local optimization problem in GAs and to examine the relationships among performance of GAs, genetic parameters, and the accuracy of LCLU classification. The relationships among them are very complex. Based on the results of the experiment, some conclusions can be drawn:

- When genetic parameters are set improperly (*e.g.*, too small number of generations, small population size, and low crossover rate), the premature convergence / local optimization occurs, and premature convergence / local optimization will lead to poor LCLU classification accuracy. Similarly, the GA-based LCLU classification cannot perform well when the mutation rate and generation gap is too high.
- In order to improve the use of GAs in LCLU classification, genetic parameters should be set using moderate values. The range of values should be as follows: generation 2000 - 5000, population 1000, crossover rate 69% - 99%, mutation rate 0.1% - 0.5%, and generation gap 25% - 50%.

CHAPTER 5 IMPACT OF IMAGE PARAMETERS ON LAND-COVER AND LAND-USE CLASSIFICATION ACCURACY

5.1. Experiment #2 Design

The accuracy of any LCLU classification is affected by a number of image parameters. The experiment designed in this chapter is used to examine the impact of image parameters on LCLU classification in the Tickfaw watershed. Figure 5.1 shows the procedure of applying GAs, expert systems, and various layer combinations in the LCLU classification. A Landsat image acquired on Feburary 6, 2000 (Figure 3.2), the resultant image of principal components analysis (PCA), the resultant image of tasselled cap transformation (TCT), the resultant image of texture analysis (TA), the resultant image of iterative self-organizing data analysis (ISODATA), normalized difference vegetation index (NDVI), normalized land-water index (NLWI), digital elevation model (DEM), soil data, and their various combinations were used in this experiment.

During the data preparation phase, an ISODATA containing 50 classes was created through ERDAS Imagine. A total of 50000 random samples was taken using the procedure described in Section 4.1 of Chapter 4. An ASCII file that contains the coordinates for all random samples was also created. This coordinate file was used to create both training data and testing data from each band (1 - 5, and 7) and other data layers (such as PCA components or DEM) (Figure 5.2a). After the random sampling, the data are converted into an attribute-relation file format (ARFF) file (Figure 5.2b).

Based on the proposed questions related to the research objectives, six types of image parameters were tested in this experiment: spatial resolution, training data size, different transformed data, different index data, different GIS data, and different data combinations. In order to investigate the impact of different data on GA-based LCLU classification, only six bands of Landsat image and one type of ancillary data (*e.g.*, DEM or first 6 PCA components)

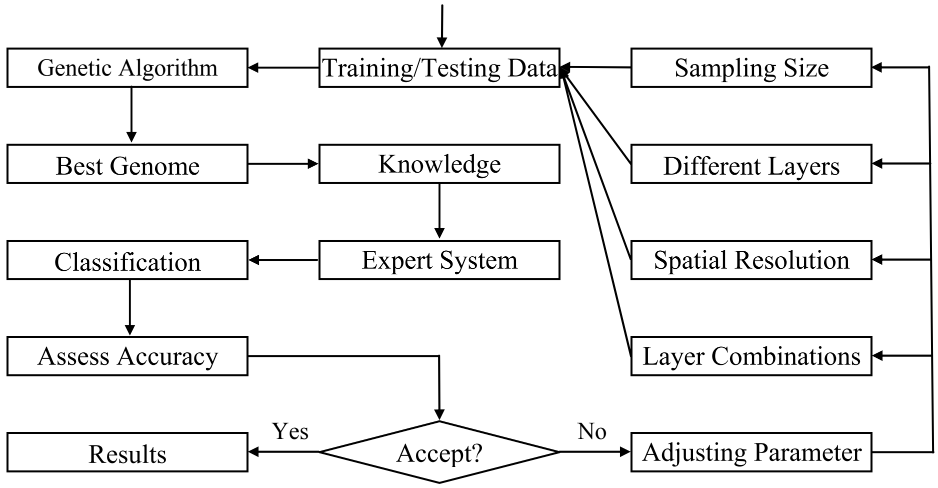
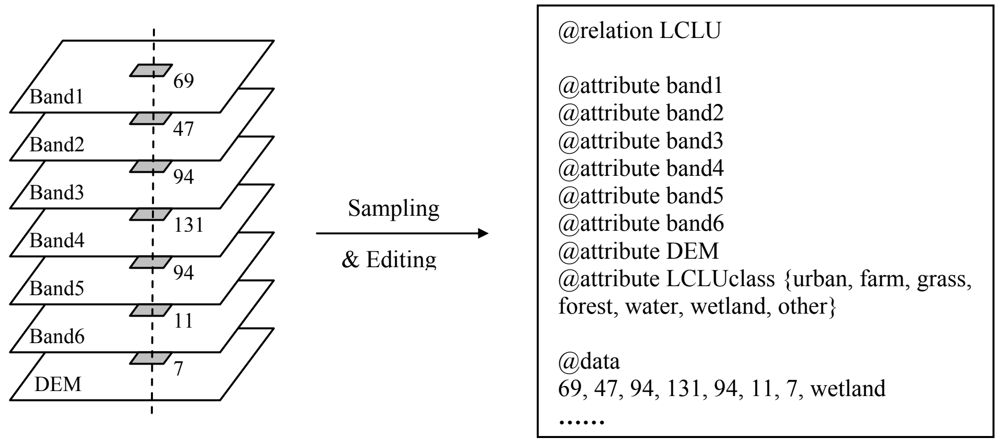


Figure 5.1. The workflow for optimizing image parameters configuration in genetic algorithm-based land-cover and land-use classification



(a) (b)

Figure 5.2. (a) Randomly sampled training data and (b) Edited data file (during optimizing image parameters)

were used at a time. After testing all ancillary data, various combinations of ancillary data were also examined. Each test generated a set of genomes. The best genome was represented by a decision tree, and it was translated into a set of classification rules (Table 5.1). The classification rules were then implemented to encode an expert system using ERDAS Imagine 9.0. During the experiment, the impacts of various training / testing data sizes, various spatial resolutions, and various data combinations were examined.

In order to investigate the impacts of different data combinations on the performance of GA-based LCLU classification, four data combinations were created. The first data set includes 5 layers: Landsat ETM image 2000, Bands 1 - 4, and Band 7. The second data set includes 12 layers: all data layers in the first data set, four PCA components (1, 2, 3, and 4), and three TCT bands (1, 3, and 4). The third data set includes 16 layers: all data layer in the second data set, three TA bands (1, 2, and 4), and ISODATA. The fourth data set includes 20 layers: all data layers in the third data set, NDVI, NLWI, DEM, and soil data. Those data layers were selected by using the GA-based attributes selection function in Weka 3.5 (Witten and Frank, 2005).

This study only focuses on the impacts of image parameters on GA-based LCLU classification. The default setting for carrying out this experiment was: training / testing data size 20000 / 10000, spatial resolution 30 meters, number of generations 5000, population size 1000, crossover rate 99%, mutation rate 0.5%, and generation gap 50%. Again, when examining the impact of a data layer (*e.g.*, DEM), the remaining parameters (*e.g.*, spatial resolution and data size) remained at the default values (Table 5.2 and Table 5.3). The accuracy assessment of LCLU classification was performed using the approach described in the previous chapter.

5.2. Results

In general, a population will converge when most individuals in the population are identical; in other words, the diversity of population is minimized (Louis and Rawlins, 1993).

Table 5.1. Translating a portion of decision tree into classification rules
(during optimizing image parameters configuration)

<table>
<tr><th>Decision Tree</th><th>Classification Rules</th></tr>
<tr><td>if band4 <= 131 then
| if band5 <= 94 then
| | if band3 <= 94 then
| | | if band1 <= 69 then
| | | | if band2 <= 47 then
| | | | | if band 6 <= 11 then
| | | | | | if DEM <= 7 then
| | | | | | | - wetland
| | | | | | + - forest
| | | | | + - grass
| | | | + - urban
| | | + - urban
| | + - other
| + - farm
|……</td><td>Rule 1. if band4 <= 131, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, band6 <= 11, and DEM <= 7, then class = wetland;
Rule 2. if band4 <= 131, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, band6 <= 11, and DEM > 7, then class = forest;
Rule 3. if band4 <= 131, band5 <= 94, band3 <= 94, band1 <= 69, band2 <= 47, and band6 > 11, then class = grass;
Rule 4. if band4 <= 131, band5 <= 94, band3 <= 94, band1 <= 69, and band2 > 47, then class = urban;
Rule 5. if band4 <= 131, band5 <= 94, band3 <= 94, and band1 > 69, then class = urban;
Rule 6. if band4 <= 131, band5 <= 94, and band3 > 94, then class = other;
Rule 7. if band4 <= 131, and band5 > 94, then class = farm.
……</td></tr>
</table>

Table 5.2. Relationship between individual image characteristics and classification accuracy

Data	Case ID*	Size, Rate, or Layers	Best Genome Fitness	Average Genome Fitness	Best Genome Size	Training Data Classification Accuracy	Testing Data Classification Accuracy
Spectral Only	SO	30m resolution, 20000/10000 Training/testing samples, 6 or 6+1 layers, and optimum GA parameter configuration*	0.6411	0.6408	135	80.15%	78.78%
+ PCA	K2		0.6404	0.6398	161	80.13%	79.12%
+ TCT	K3		0.6551	0.6547	131	81.01%	80.27%
+ TA	K4		0.6782	0.6777	181	82.49%	81.30%
+ ISO	K5		0.6579	0.6576	163	81.22%	79.99%
+ NDVI	K6		0.6370	0.6367	167	79.93%	78.44%
+ Water Index	K7		0.6722	0.6718	115	82.04%	80.98%
+ DEM	K8		0.7531	0.7531	161	86.92%	85.92%
+ Soil data	K9		0.7276	0.7273	157	85.41%	84.42%

* Optimum GA parameter configuration: number of generations: 2000, population size: 1000, crossover rate: 99%, mutation rate: 0.5%, and generation gap: 50%.

Table 5.3. Relationship between image characteristics combination and classification accuracy

Data	Case ID**	Size, Rate, or Layers	Best Genome Fitness	Average Genome Fitness	Best Genome Size	Run Time (min.)	Training Data Classification Accuracy	Testing Data Classification Accuracy
Data Comb.	A1	5	0.6182	0.6180	153	1934	78.72%	77.52%
	A2	12	0.6522	0.6511	213	1710	80.94%	79.97%
	A3	16	0.7085	0.7079	167	2005	84.29%	83.51%
	A4	20	0.7790	0.7783	165	1740	88.38%	87.47%
Training / esting Data Size	D1	1000 / 500	0.8255	0.8243	231	125	91.10%	84.00%
	D2	3000 / 1500	0.7731	0.7725	175	360	88.06%	84.72%
	D3	5000 / 2500	0.7688	0.7671	201	611	87.86%	85.73%
	D4	10000 / 5000	0.7691	0.7676	167	1200	87.82%	87.06%
	D5	20000 / 10000	0.7790	0.7783	165	1740	88.38%	87.47%
	D6	30000 / 15000	0.7750	0.7745	191	3690	88.29%	88.20%
	D7	50000 / 25000	0.7734	0.7723	183	4080	88.09%	87.82%
Spatial Resolution	S1	7.5m	0.7599	0.7594	147	1739	87.27%	86.17%
	S2	15m	0.7653	0.7644	177	1656	87.62%	86.88%
	S3	30m	0.7790	0.7783	165	1740	88.38%	87.47%
	S4	60m	0.7934	0.7970	159	1893	89.41%	88.79%
	S5	120m	0.7668	0.7656	179	1750	87.71%	86.92%
Optimum Sitting	OS	Genetic Parameters	5000 generations, 1000 individuals, 99% crossover, 0.5% mutation, 50% generation gap					
		Data Characteristic	20 layers data combination, 20000/10000 training / testing samples, 30 m spatial resolution,					
		Results	0.8309	0.8300	203	--	91.34%	90.67%
** Condition: number of generation: 5000, population size: 1000, crossover rate: 99%, generation gap: 50%, mutation rate: 0.5%, training / testing data size: 20000 / 10000, spatial resolution: 30 m, data combination: 5-20 layers.								

Different populations converge to different fitness levels. A population with high diversity will eventually converge to a global optimum fitness. A population with low diversity will quickly converge to a local optimum fitness. Models based on a global optimum can be applicable to most data, whereas models based on local optimum cannot. The goal of this experiment is to determine the relationship among image parameters, GA performance, and LCLU classification. Four categories of image parameters examined in the experiment were: individual data characteristics, training / testing data size, spatial resolution, and data layer combinations.

5.2.1. Individual Data Characteristics

Figures 5.3 and 5.4 illustrate the relationship between the accuracy of GA-based LCLU classification and data layer. After combining six bands of spectral data with another ancillary data (*e.g.*, DEM or PCA components), the performance of GA-based LCLU classification was improved, and different ancillary data improved GA-based classification differently. DEM, soil data, texture data, and land-water interface data can significantly improve GA's performance. Before adding those data, the classification accuracy of training / testing data was 79.93% / 78.44%. After adding those types of data, the classification accuracy of training / testing data became 86.92% / 85.92%, 85.41% / 84.42%, 82.49% / 81.30%, and 82.04% / 80.98% respectively. The rate of improvement for training / testing data classification is 2.64-8.75% / 3.14-9.54%. The results also indicated that PCA only slightly improved the performance of GA-based classification. These results only reflect the improvement in terms of model development or knowledge discovery, and they were based on the training / testing data only. Whether they improve the real data classification or not will be examined in the following sections.

5.2.2. Training / Testing Data Size

In a knowledge-based LCLU study, the quality and quantity of sampling data are always an important issue. It could affect the quality of knowledge. Generally, a training data set with

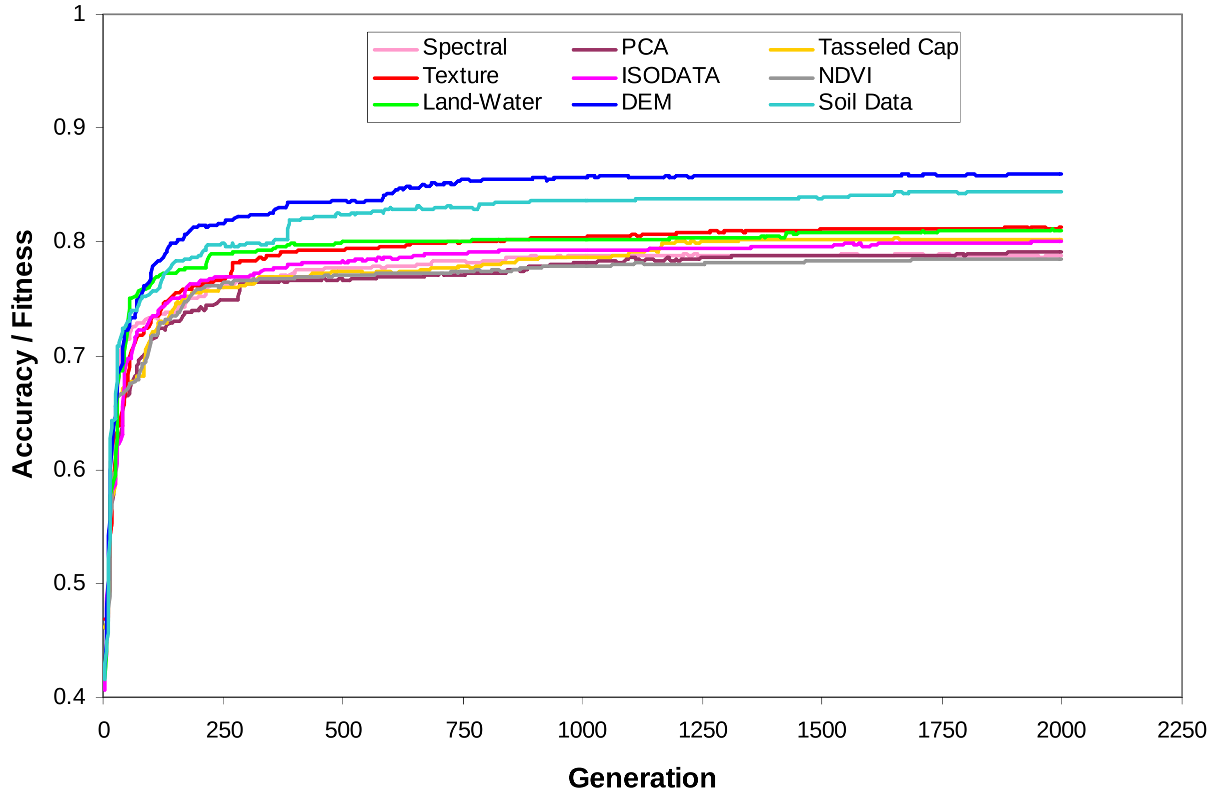


Figure 5.3. Relationship between accuracy / fitness and data layers

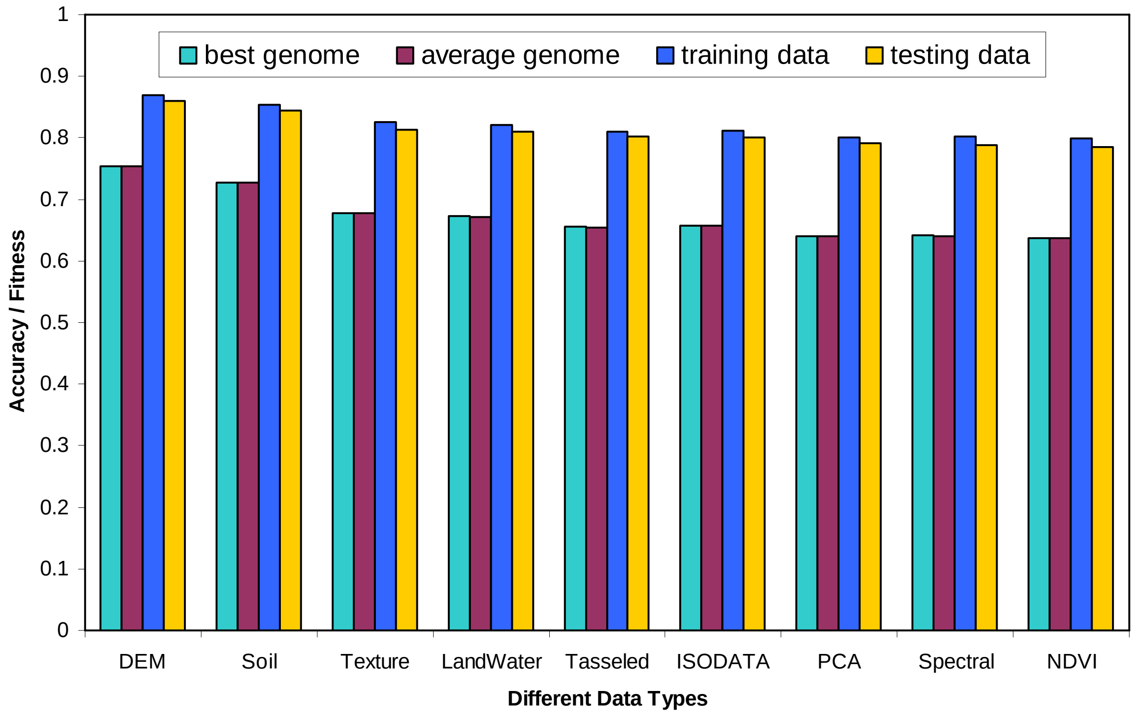


Figure 5.4. Comparing accuracy / fitness of population
when data layer changes

high sampling rate has small rate of unknown data, and it can better represent the original data. However, a large training data size usually leads to high complexity, more noise, and more time to process. Although GAs were proposed as a random search and global optimization technique, the search space is still based on the training data. The knowledge, if derived from the training dataset that poorly represent the original data, will not be useful for analyzing the pattern of original data.

Figures 5.5 and 5.6 show the relationships among training data size, fitness, and classification accuracy. During a 5000-generation evolution, the best/average fitness first increased dramatically, then increased slowly, and stabilized around 0.77 while the training/testing data size increased (Table 5.3). Population, derived from small training data sine (*e.g.* 1000 pixels), had the highest best/average fitness (0.8255 / 0.8243) and training data classification accuracy (91.10%), but it had the smallest testing data classification accuracy (84.00%) except the optimum case. This is a typical phenomenon of local optimum. Due to a low sampling rate, the diversity of population is small, which leads to premature convergence / local optimum. The knowledge better fitting the training data may not better fit the testing data or real world data. The results also indicate that large training data size may not always improve the performance of GAs in terms of fitness and classification accuracy (*e.g.*, the comparison among training data with 30000 and 50000 pixels). One possible reason is that large training data may be more complex than small training data, and it needs more time to develop a high fitness population. Given the same number of generations (*e.g.*, 500), it is difficult for a large training data-based population to develop a better fitness than small-size training data. However, population based on a large training data size (*e.g.*, 20000), even with lower training data classification accuracy, still has higher testing data classification accuracy than that of small-size training data (*e.g.*, 1000).

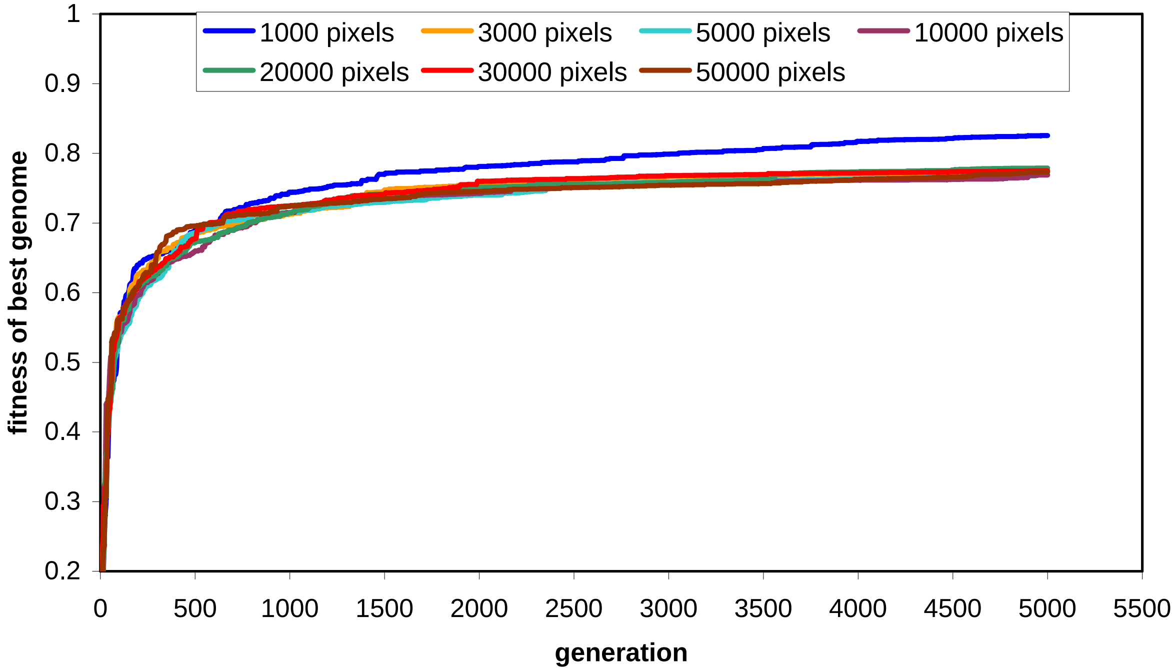


Figure 5.5. Relationship between accuracy / fitness and sampling size

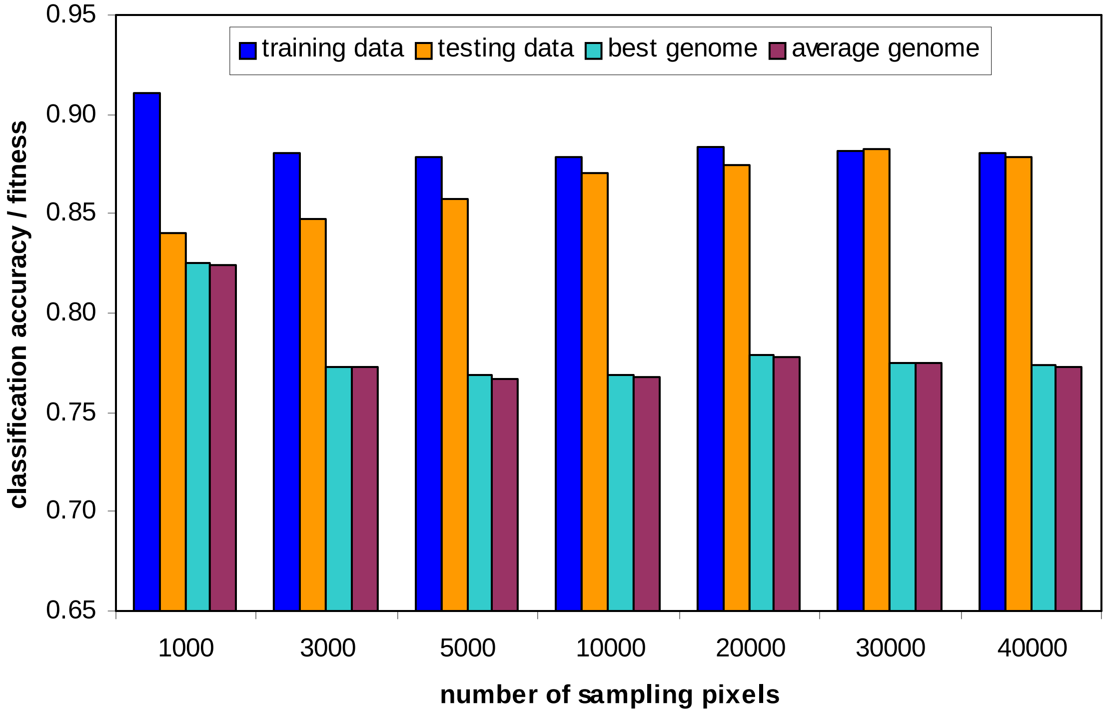


Figure 5.6. Comparing the accuracy / fitness of population
when sampling size changes

5.2.3. Spatial Resolution

Five different data sets were used in the experiment for examining the impact of spatial resolutions. As they have been described in the previous section, those spatial resolutions were: 7.5 m, 15 m, 30 m, 60 m, and 120 m. They were created by re-sampling using the nearest neighbor algorithm. It is commonly recognized that there is a relationship between the spatial resolution of a satellite image and the classification accuracy of LCLU (Cushnie, 1987). The advantage of a high spatial resolution image is obvious when being visually compared with low spatial resolution image. Unfortunately, the refinement of spatial resolution from low resolution to high resolution may not always improve classification accuracy (Irons *et al.*, 1985). Although increasing spatial resolution tends to improve classification accuracy through decreasing the proportion of mixed pixels, it also tends to depress classification accuracy through increasing spectral variability and decreasing separability (Irons *et al.*, 1985). The dominant trends of improvement or mis-classification depends on the classification scheme and the average field size within a scene (Yang and Lo, 2002). The improved spatial resolution of the original Landsat TM image did not greatly improve the classification accuracy. Because improved spatial resolution can lead to an increase not only in the inter-class variability but also the intra-class variability, both types of variability can lead to poor classification accuracy (Yang and Lo, 2002).

Figure 5.7 shows the evolution process of a population derived from different training data sets. Unlike the refined or degraded data, the original data-based population seems to converge smoothly. A possible reason is that population based on the refined or degraded data may be dominated by high-fitness individuals during the early stage of evolution. Figure 5.8 compares the best/average fitness and classification accuracy of different training/testing data set. It indicates that there are both improvements and hindrances on classification accuracy in both cases of refined and degraded spatial resolution. This result confirms that the effect of spatial resolution on classification accuracy is very complex, and it cannot be simply described as improvement or hindrance (Irons *et al.*, 1985; Cushnie, 1987).

The training/testing data can be recognized as the environment of population. The complexity of the environment increases when spectral variability increases, and the complexity of environment decreases when the number of mixed pixels decreases. To fit a complex environment, the population requires a long time for evolution.

Higher spatial resolution may provide a chance to glean more precise information. Therefore training/testing data developed from higher resolution data usually leads to high fitness. On the other hand, it may lead to poor fitness as well, because if the sampling size remains constant, un-sampled / unknown area will increase. For example, at the same sampling size level, 15 m resolution data has lower training/testing fitness, 30 m resolution data has higher training/testing fitness. It should be noted that higher training / testing fitness do not ensure higher classification accuracy. For example, when the sampling rate is low, the relative noise data rate is lower. It is easy to implement a model to capture most training data and testing data. However, since the relative un-sampled / unknown area is larger, the model cannot capture most data characteristics in a real case.

5.2.4. Data Layers Combinations

Four different data combinations were used in the experiment. As they have been described in a previous section, those data combinations included 5 layers, 12 layers, 16 layers, and 20 layers (Figure 5.9). Although the population in GAs is initialized randomly (Goldberg, 1989), its development is still based on the training data. The training data can be recognized as the environment of the population. Generally, a population that lives in a complex environment will become complex. Otherwise, it will be eliminated through the natural selection. According

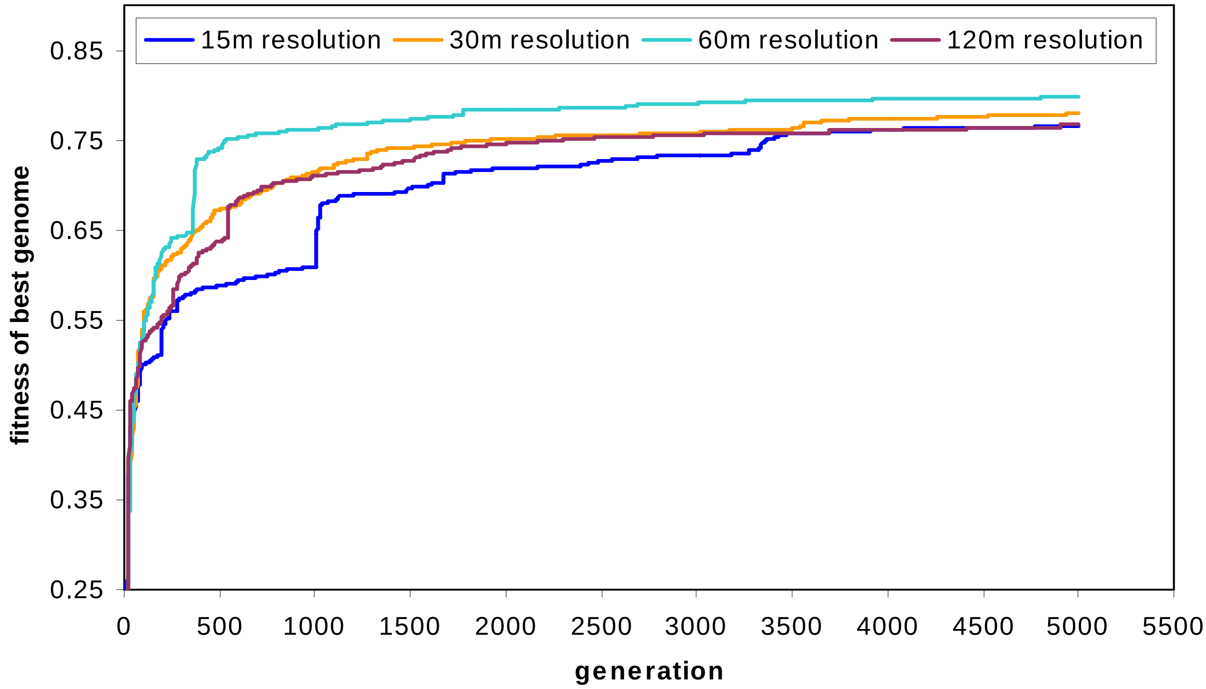


Figure 5.7. Relationship between accuracy / fitness and spatial resolution

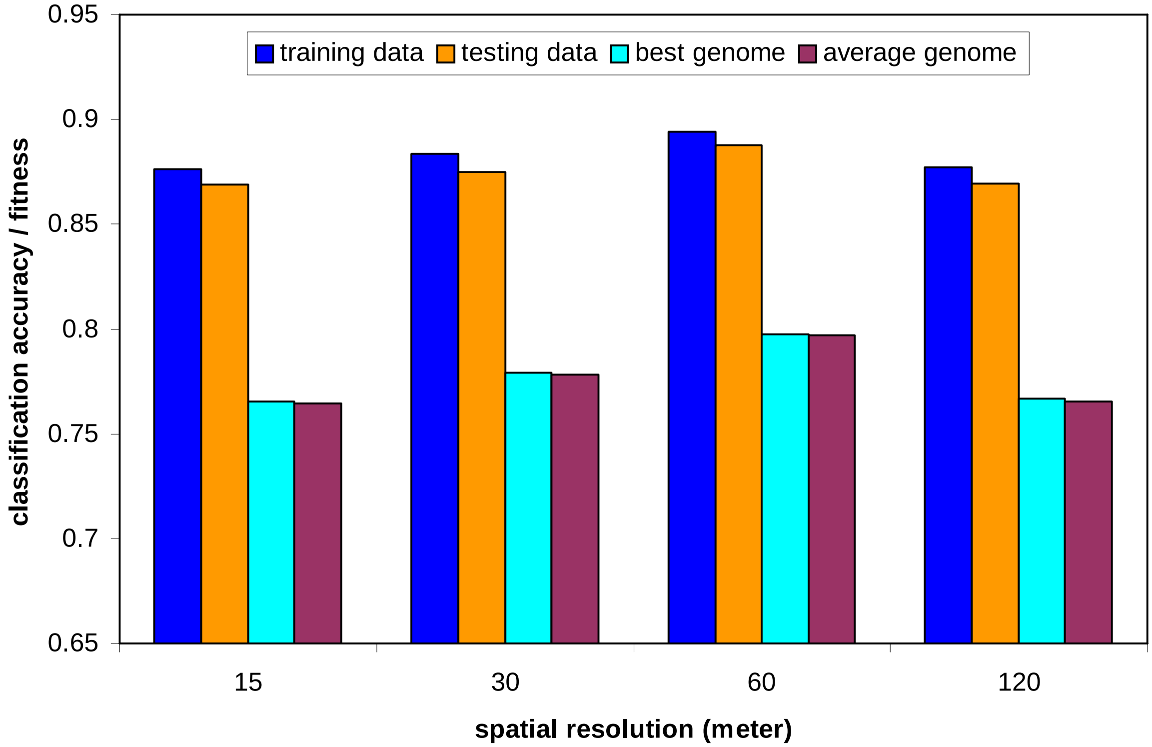


Figure 5.8. Comparing the accuracy / fitness of population
when spatial resolution changes

to the research done by Wang and Cao (2002), the more complex the genome is, the larger space it can search, and the higher fitness/classification accuracy it can reach. The research done by geographers has proved that GIS data (such as soil data) and other data (such as DEM) can benefit knowledge-based LCLU classification (Huang and Jenson, 1997).

Figure 5.9 shows the evolutionary process of population derived from different training data sets. Figure 5.10 compares the best/average fitness and training/testing data classification accuracy derived from different training/testing data. The more layers the training data contains, the higher fitness/classification accuracy the population can achieve. Many scholars believe that there is a connection between the complexity of population and the convergence rate, and complex population usually converged slowly (Papagelis and Kalles, 2001; Wang and Cao, 2002). But, a comparison of the performance of genetic algorithms with different training/testing data suggests differently (Table 5.3). Population – based on first, second, third, and fourth training data set – required 1934, 171, 2005, and 1740 minutes, respectively, to complete 5000-generation evolution.

5.2.5. Land-Cover and Land-Use Classification

After discovering the spatial spectral knowledge, a set of classification experts were encoded and used for the LCLU classification. Table 5.4 and Figure 5.11 demonstrate the results of classification. On the overall level, the models with high fitness can achieve high LCLU classification accuracy. For example, Case K9 has high best / average genome fitness (0.7276 / 0.7273), high training / testing data classification accuracy (85.41% / 84.42%), and high LCLU classification accuracy (88.30%). The models with low fitness may achieve low LCLU classification accuracy. For example, Case K5 has low best / average genome fitness (0.6579 / 0.6576), low training / testing data classification accuracy (81.22% / 79.90%), and low LCLU classification accuracy (75.60%). However, some cases do not follow this pattern. Case K2 has

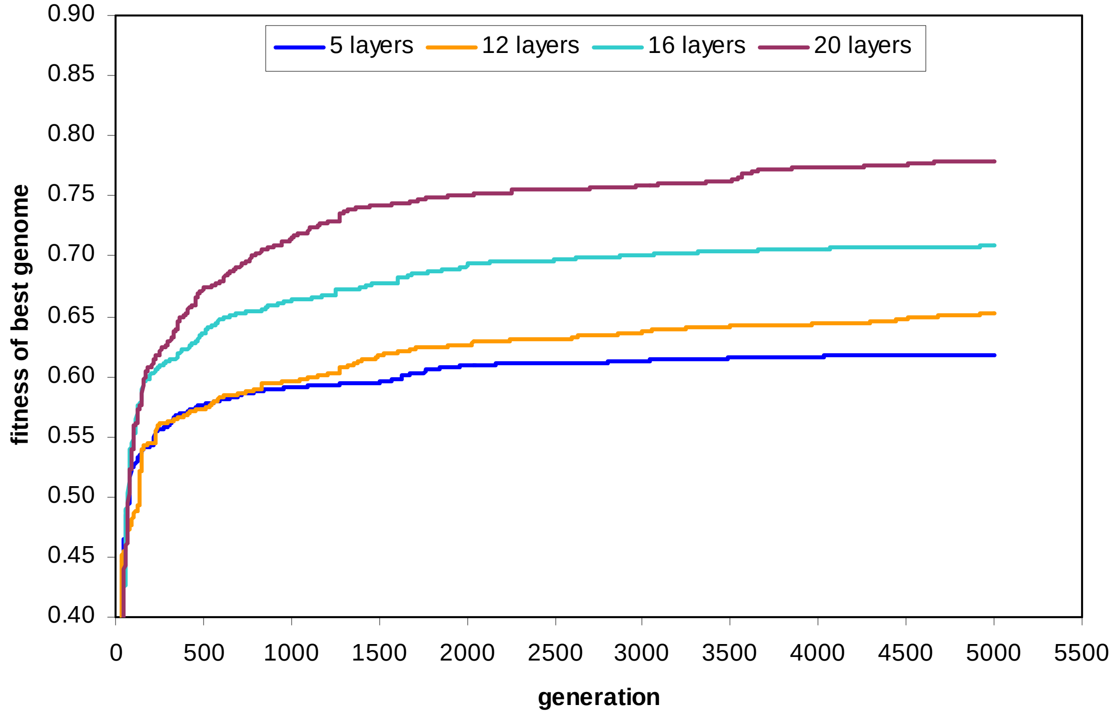


Figure 5.9. Relationship between accuracy / fitness and layer combination

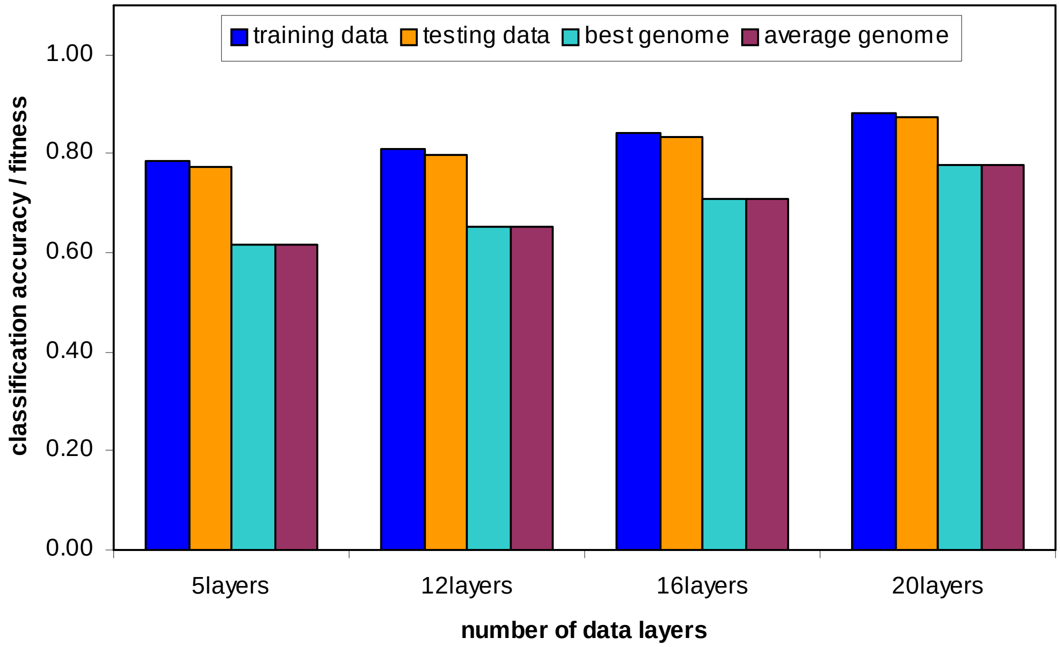


Figure 5.10. Comparing the accuracy / fitness of population when layer combination changes

low best / average genome fitness (0.6404 / 0.6398), low training / testing data classification accuracy (80.13% / 79.12%), and high LCLU classification accuracy (86.10%). Case D1 has high best / average genome fitness (0.8255 / 0.8243), high training / testing data classification accuracy (91.10% / 84.00%), and low LCLU classification accuracy (75.60%).

On the class level, most models have good producer accuracy and user accuracy. For example, both Case K3 and Case K9 have high producer accuracy and user accuracy on all features except land feature GRASS. In this research, GRASS has the lowest producer accuracy and user accuracy in most cases (Table 5.4).

5.3. Discussion

In the real world, the individuals, from the same or a different population, have similar / different genome and different strategies to fit the different environment. In Chapter 4, the genomes were developed by using only the Landsat image (Bands 1-5, and 7). They can be considered as individuals from the same population. Different from the experiment in Chapter 4, this chapter developed the genomes by using different data layers and layer combinations (*e.g.*, Landsat image + DEM, or Landsat image + soil data), and they can be considered as different individuals from different population. Therefore, the impact of image parameters on GA-based LCLU classification was categorized into at least five aspects: 1) the impact of an individual data layer; 2) the impact of a combination of data layers; 3) the impact of training/testing data size; 4) the impact of spatial resolution; and 5) the recommendation of configuration of image parameters.

5.3.1. Impact of Individual Data Layer

Most existing LCLU classification applications are spectrally-based. The combination of spectral and spatial classification is especially valuable for accurate LCLU classification in areas with complex landscapes (Lu and Weng, 2007). This research integrates GIS information

Table 5.4. Classification accuracy assessment for image parameter configurations

Designed Experiments		**Spectral Only ***	**+DEM***	**+Soil Data***	**+Texture Analy. ***	**+Land-Water ***	**+ISO Data***	**+PCA***
Case ID		SO	K8	K9	K4	K7	K5	K2
Urban	Prod. Accu.	80.68%	88.07%	87.50%	86.36%	91.48%	88.64%	86.93%
	User Accu.	88.20%	85.64%	88.00%	68.47%	68.80%	46.90%	77.66%
Farm	Prod. Accu.	92.46%	93.47%	93.97%	90.95%	87.94%	89.95%	92.96%
	User Accu.	78.30%	77.50%	82.74%	76.69%	82.55%	82.49%	76.65%
Grass	Prod. Accu.	38.04%	39.13%	48.91%	20.65%	19.57%	35.87%	22.83%
	User Accu.	54.69%	76.60%	61.64%	63.33%	62.07%	61.11%	77.78%
Forest	Prod. Accu.	92.89%	92.42%	93.36%	91.00%	93.36%	41.23%	94.79%
	User Accu.	85.96%	91.55%	92.06%	91.43%	91.20%	91.58%	88.89%
Water	Prod. Accu.	87.88%	89.39%	93.94%	93.94%	96.97%	96.48%	96.97%
	User Accu.	100.00%	98.33%	98.41%	98.41%	98.46%	97.01%	100.00%
Wetland	Prod. Accu.	90.64%	95.74%	92.77%	93.62%	92.34%	93.19%	93.62%
	User Accu.	94.25%	94.94%	97.32%	97.35%	96.44%	95.63%	97.78%
Other	Prod. Accu.	90.48%	90.48%	95.24%	61.90%	80.95%	80.95%	85.71%
	User Accu.	67.86%	86.36%	80.00%	100.00%	89.47%	89.47%	90.00%
Overall Accuracy		84.70%	87.50%	88.30%	83.90%	84.90%	75.60%	86.10%
Overall Kappa		0.8119	0.8459	0.8564	0.8011	0.8138	0.7014	0.8283

Table 5.4. (Continued)

Designed Experiments		+NDVI*	+Tasseled Cap*	Training Data**	Spatial Resolut**	Data Comb.**	Optimum Sitting **
Case ID		K6	K3	D1	S1	A1	OS
Urban	Prod. Accu.	82.39%	85.80%	88.07%	94.32%	81.82%	93.18%
	User Accu.	83.82%	69.27%	75.24%	74.44%	70.94%	87.70%
Farm	Prod. Accu.	93.47%	91.96%	90.45%	92.46%	91.46%	90.45%
	User Accu.	78.81%	78.21%	85.31%	85.19%	80.89%	84.51%
Grass	Prod. Accu.	33.70%	10.87%	36.96%	52.17%	23.91%	56.52%
	User Accu.	55.36%	100.00%	60.71%	64.86%	64.71%	73.24%
Forest	Prod. Accu.	92.42%	93.36%	55.45%	76.30%	93.36%	92.89%
	User Accu.	85.53%	89.14%	92.86%	93.06%	87.95%	91.59%
Water	Prod. Accu.	95.45%	96.97%	75.76%	98.48%	93.94%	89.39%
	User Accu.	100.00%	84.21%	100.00%	95.59%	100.00%	98.33%
Wetland	Prod. Accu.	90.64%	89.79%	92.77%	92.77%	93.62%	95.32%
	User Accu.	95.09%	96.35%	96.89%	96.04%	96.92%	96.97%
Other	Prod. Accu.	85.71%	95.24%	71.43%	90.48%	90.48%	95.24%
	User Accu.	90.00%	90.91%	68.18%	100.00%	79.17%	100.00%
Overall Accuracy		85.10%	83.60%	76.90%	86.10%	84.60%	89.50%
Overall Kappa		0.8165	0.7977	0.7228	0.8297	0.8101	0.8711

* Condition: generation amount: 2000, population size: 1000, crossover rate: 99%, mutation rate: 0.5%, generation gap: 50%, training / testing data size: 20000 / 10000, spatial resolution: 30 m, data combination: 6 or 6+1 layers.
** Condition: generation amount: 5000, population size: 1000, crossover rate: 99%, mutation rate: 0.5%, generation gap: 50%, training / testing data size: 20000 / 10000 (except case D1), spatial resolution: 30 m, data layer combination: 5 or 20 layers.

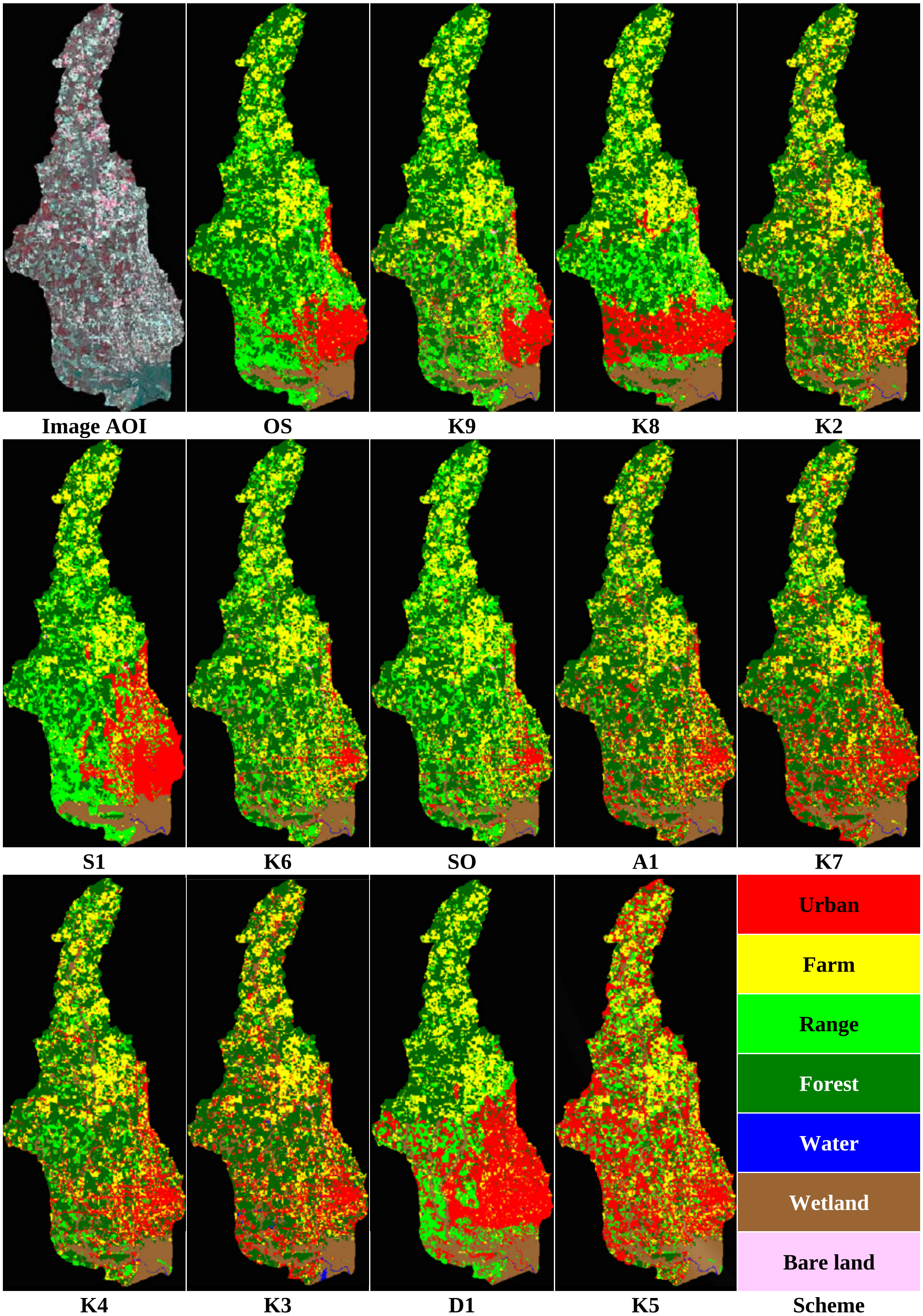


Figure 5.11. The results of case study on genetic algorithms-based land-cover and land-use classification

contained in a DEM and soil data into Landsat image classification. The results indicated that both DEM and soil data helped to increase the classification accuracy by 2.8% and 3.7%, respectively. Usually a knowledge-based LCLU classification with Landsat image and GIS data yielded higher accuracy than a standard spectrally-based classification (Bolstad and Lillesand, 1992). The spatial distribution of natural and man-made features is closely related to DEM and soil data. For example, hydrophytes usually live in the low DEM area, and high-population density areas normally have a higher DEM. Also different soil types usually support different vegetation or different agricultural activities. Therefore, DEM and soil data can improve LCLU classification.

In addition to GIS data, the study used Mean Euclidean Distance (MED) based texture analysis to attain spatial information. Texture analysis measures the spatial distribution of ground radiance level variations, which is closely related to the structures of a satellite image, and it can be used to characterize the morphology of study area. The research attempted to integrate spatial information derived from texture analysis into the LCLU classification. However, the accuracy was only improved in the phase of model development (Table 5.2). In the phase of LCLU classification, the classification accuracy was not improved at all. Possible reasons are: 1) the large portion of Tickfaw watershed is covered by forest, agricultural land, and grass land; so its morphological characteristics are relatively simple, 2) texture analysis is usually best-suited for high-resolution imagery, and the data used in this study are only at 30 m resolution.

Two band-ratioing techniques, namely normalized difference vegetation index (NDVI) and normalized land water interface (NLWI), were used in the study. NDVI is a commonly-used vegetation index. Usually, heavily vegetated areas display high positive values, whereas high-density residential areas have low NDVI values. The results demonstrate that after incorporating NDVI in the experiment, the classification accuracy of farm, grass, and forest was not improved, but non-vegetated features, such as water and bare land was significantly improved. After using NDVI, there is still a mixed class situation among farm, grass, and forest. The results show that the fitness and the classification accuracy of training / testing data slightly decreases (Table 4.2), but the accuracy of Landsat image classification increases (Table 5.2). These results indicate that: 1) the NDVI can distinguish vegetated features and non-vegetated features, but it is not a good approach for separating farm, grass, and forest, 2) different individuals (model) from different population (a set of models) fit the same environment (Landsat image) differently. A model may not fit a data set, but it may fit another data set very well.

In addition to NDVI, another band ratio, namely NLWI was used in the study. NLWI is the index used to separate water body or high-wetness feature from dry land features (Gao, 1996). Although the integration of NLWI and Landsat imagery only slightly improved the overall accuracy of LCLU classification, the producer accuracy of water feature was significantly improved. This indicates that NLWI can extract water features from non-water features efficiently. However, water features only occupy a small portion of study area, so the improvement of overall accuracy was not significant.

Two transformation techniques, namely principal component analysis (PCA) and tasseled cap transformation (TCT), were used in the study. PCA is one of the most important linear transformation techniques (Coppin *et al*., 2004), and it is usually used to reduce data redundancy between bands and highlight different information in the derived components (Lu *et al*., 2004a). After incorporating principal components into the classification procedure, the classification accuracy of the testing data and Landsat image was improved. The reason for this improvement is that PCA not only eliminates noise and redundancy but also emphasizes different information. Another transformation technique in the study is the TCT. It can also be used to reduce data redundancy. One of its advantages over PCA is that its transform coefficients are independent of the image scenes, while PCA is dependent on the image scenes (Lu *et al*., 2004a). It generates the coefficients of brightness, greenness, and wetness. The study shows that TCT bands can improve either producer accuracy or user accuracy. However, the overall accuracy was not improved. Possible reasons are the following: 1) the brightness coefficients cannot separate urban and bare land efficiently, 2) the greenness coefficients cannot separate agricultural land, range, and forest, and 3) the greenness coefficients cannot separate water and wetland efficiently.

Finally, a clustering techniques technique, namely iterative self-organizing data analysis (ISODATA), was also used in the study. This technique is implemented by recursively migrating a set of cluster means (centers) using a "closest distance to mean" approach until the locations of the cluster means are unchanged (Tso and Mather, 2001). In this study, the number of clusters to be produced was specified as 50. After integrating the ISODATA and Landsat image into the procedure of knowledge discovery, the model was improved in terms of fitness and training / testing data classification accuracy. But the accuracy of real data classification was not improved. There may be several reasons for this case: 1) although training / testing data was taken from the Landsat image data, the model (or individual) which well fits training / testing data (partial environment) may not fit the entire AOI of the image data (entire environment), 2) ISODATA generates clusters based on pixel value, and pixels with same value may represent different land features, and 3) the capability of ISODATA is very limited, and it can not separate mixed classes (Huang and Jensen, 1997).

5.3.2. Impact of Data Layer Combinations

The accuracy of the LCLU classification can be improved by using: 1) different algorithms or techniques, 2) different remotely-sensed images, 3) different image features, and 4) different ancillary data (Lu and Weng, 2007). The results shown in Section 5.2.1 indicate that different image features and different ancillary data play various roles in the improvement of

LCLU classification accuracy. The combinations of those image features and ancillary data also display very significant results. The major understanding developed from the result is that more useful data can benefit the classification (Figures 5.9 and 5.10).

Both Figures 5.9 and 5.10 show that the more layers the training data contains, the higher the fitness / classification accuracy the population can achieve. In addition, many scholars believe that there is a connection between the complexity and the convergence rate, and complex populations converged slower than simple ones (Papagelis and Kalles, 2001; Wang and Cao, 2002). However, the comparison of the performance of GAs with different training / testing data size suggested otherwise (Table 5.3). Population based on A1, A2, A3, and A4 training data sets (from simple to complex) required 1934, 1710, 2005, and 1740 minutes to complete the 5000 generation evolution, respectively. It suggests that population complexity is not directly associated with evolutionary time.

5.3.3. Impact of Training / Testing Data Size

The relationship between training / testing data size and the performance of GA-based LCLU classification is complex. The training / testing data with the smallest size has the highest training data classification accuracy and the lowest testing data classification accuracy (*e.g.*, case D1) (Table 5.3), and the knowledge derived from such data can lead to poor LCLU classification (Table 5.4). This is because data with a lower sampling rate has higher rate of unknown point, and population derived from such data has lower biodiversity. Therefore, the knowledge cannot be used to classify LCLU accurately.

The training / testing data with the larger size has higher training data classification accuracy and higher testing data classification accuracy (*e.g.*, Case D5) (Table 5.3), and the knowledge derived from such data set can lead to more accurate LCLU classification (*e.g.*, case SO) (Table 5.4). This is because data with higher sampling rate has lower rate of unknown point, and population derived from such data has higher biodiversity. Therefore, the knowledge can be used to classify LCLU accurately.

5.3.4. Impact of Spatial Resolution

Spatial resolution determines the level of spatial detail that can be observed on the Earth's surface (Lu and Weng, 2007). Spatial resolution is one of the most important image features that affects LCLU classification (Chen *et al*., 2004; Atkinson and Curran, 1997; Atkinson and Aplin, 2004). It determines the level of spatial detail that can be observed on the Earth's surface (Lu and Weng, 2007). Although high spatial-resolution imagery can provide a wealth of detailed information about the ground, the classification results are not as promising as expected. Empirical studies have shown that increasing spatial resolution does not necessarily improve classification accuracy (Hsieh *et al*., 2001; Landgrebe *et al*., 1977; Latty *et al*., 1985; Williams *et al*., 1984; Toll, 1985). Therefore, it is necessary to select a proper spatial resolution for the particular application (Atkinson and Curran, 1997). In this study, five different spatial resolutions were designed and implemented in this experiment (Table 5.3). All of them were tested in the phase of knowledge discovery, and two of them were tested in the phase of LCLU classification. Based on the results from the phase of knowledge discovery (Table 5.3), when spatial resolution decreased from 7.5 m to 60 m, the fitness of best / average genome and the classification accuracy of training / testing data increased. When spatial resolution decreased from 60 m to 120 m, the fitness of best / average genome and the classification accuracy of training / testing data decreased. These results again showed that refining spatial resolution did not necessarily improve classification accuracy (Irons *et al*., 1985), as it can lead to an increase not only in the inter-class variability but also in the intra-class variability (Williams *et al*., 1984; Irons *et al.*, 1985; Haack *et al.*, 1987). The increased variability decreases the statistical separability of land features classes. This decreased separability tends to depress classification accuracies when pixel-based classification approaches are used. The increased variability was attributed to the imaging of diverse class components by high resolution sensors. At coarser resolutions, sensors integrated the reflected spectral radiance of the various components, and classes appeared more homogeneous (Irons *et al*., 1985). Another consequence of refining spatial resolution is a decrease in the proportion of mixed pixels to pure pixels (Irons *et al*., 1985). The decreasing proportion of mixed pixels with increasing spatial resolution tends to improve classification accuracy, and therefore counteracts the consequences of increased spectral variability.

The dominant trend depends on the classification scheme and the average field size within a scene (Irons *et al*., 1985). In this study, the number of mixed pixels decreases when the resolution of the image changes from 60 m to 7.5 m, but both inter- and intra-class variability increases. The balance between mixed pixels decreasing and inter- / intra-class variability increasing will impact the accuracy of LCLU classification. The number of mixed pixels increases when the resolution of the image changed from 60 m to 120 m, which will affect the accuracy of LCLU classification. Based on the results from the phase of LCLU classification (Table 5.4), it can be concluded that the classification accuracy of high spatial-resolution imagery (7.5 m) is not always better than the classification accuracy of low spatial-resolution image (30 m).

5.3.5. Recommendation of Image Parameters Configuration

Based on the above results and discussion, two recommendations were provided in this study. The first one was based on range, and the second one was based on a particular case. The recommendation of range include: data combinations: 16-20 layers, training / testing data size: 10000/5000 – 30000/15000, spatial resolution: 30 m – 60 m, and genetic parameters will be the same optimum configuration provided in Chapter 4. The recommendation of particular case include: data combinations: 20 layers, training / testing data size: 20000/10000, spatial resolution: 30 m, and the optimum genetic parameter configuration. Table 5.4 and Figure 5.11 show the result of satellite image classification accuracy for different cases. The recommended case has highest classification accuracy for both testing data and satellite image (90.67% and 89.50%). These results suggest that the performance of genetic algorithms and LCLU classification can be improved by adjusting genetic parameters.

5.4. Hypothesis # 2 Review

The second research question as presented in Chapter 1 was how do image parameters (such as spatial resolution, training data size, different indexing data, and data combination) impact the accuracy of GA-based LCLU classification. The second research hypothesis presented in Chapter 1 was stated as follows:

Ha: Different spatial resolution, training / testing data size, data layers, and data layer combinations can increase / decrease the accuracy of GA-based LCLU classification, respectively.

Based on the experiment designed and performed in this chapter, those image parameters can impact the performance of GAs-based LCLU classification significantly. Compared to a GA-based LCLU classification using spectral data only, after adding additional data layers, the performance of GA-based LCLU classification was slightly improved. The overall classification accuracy was improved by 0.43% (PCA), 1.89% (TCT), 3.20% (TA), 1.54% (ISODATA), and 2.79% (NLWI), respectively. After adding NDVI, the accuracy of GA-based LCLU classification slightly decreased 0.43%. In the case of DEM and soil data, compared to a GA-based LCLU classification using spectral data only, after adding DEM or soil data, the performance of GA-based LCLU classification was improved significantly. The improvement rates were 9.06% (DEM) and 7.16% (soil) respectively.

In the case of training/testing data size, its impact on the performance of GA-based LCLU classification is significant. The accuracy of GA-based LCLU classification generally increases from 0.7690 to 0.8487 as the size of training/testing data increases from 1000/500 to 20000/10000.

The impacts of various image parameters on the performance of GA-based LCLU classification are different. In general, the performance can be improved by adding more ancillary data or enlarging the training data size. However, adding NDVI data or collecting too many training samples may not necessarily improve the performance of GA-based LCLU classification. The impact of spatial resolution is complex and it needs further study.

5.5. Summary

As it has been stated previously, the primary goal of this study is to examine the relationships between image parameters and the performance of GA-based LCLU classification. Based on the results of the experiment, some conclusions can be drawn:

- When training / testing data size is too small (*e.g.*, 1000 / 500), the models (or population) will be developed in a small space, so that premature convergence / local optimization occurs, which will lead to a poor classification accuracy (76.90%).
- When spatial resolution is too small / too large (*e.g.*, 7.5 m or 120 m), the variability of inter- / intra-classes and the proportion of mixed pixels will change. This means that complexity of environment changes. The performance of GA will depend on the environmental complexity. The testing data classification accuracy is 86.17% (7.5 m) and 86.92% (120 m). Both of them are smaller than 87.47% (30 m).
- Different data layers play different roles in the improvement of LCLU classification. DEM and soil data can significantly improve the classification accuracy. Other data layers can improve the classification accuracy at different levels.

- Based on this experiment, the recommended configuration of image parameters is: 16-20 layer data combination, 10000 / 5000 - 20000 / 10000 training / testing data size, 30 m - 60 m spatial resolution, and optimum genetic parameter setting as listed in Chapter 4.

CHAPTER 6 COMPARISON OF TRADITIONAL AND GA-BASED LAND-COVER AND LAND-USE CLASSIFICATION

6.1. Experiment #3 Design

The goal of the experiment is to compare unsupervised, supervised, hybrid, and GA-based approaches for LCLU classification (Figure 6.1). A special effort was made to identify the differences among them in terms of overall accuracy, kappa statistics, producer accuracy, user accuracy, omission error, and commission error. The comparison was carried out based on both overall level and class level. In order to eliminate the impact of data and focus on the algorithm itself, only the 2000 Landsat Image was used for classification in this experiment. During data preprocessing, an ISODATA image containing 160 clusters was created through the unsupervised classification in ERDAS Imagine. The procedure used 30 maximum iterations and 0.95 convergence threshold, and it output a signature file with 160 classes. The resultant data of ISODATA were then used in the unsupervised, hybrid, and GA-based LCLU classification.

For the unsupervised LCLU classification, these 160 clusters were assigned into different LCLU classes using the Raster Attribute Editor and Flicker utility of the Viewer. After assigning the clusters into different classes, they were recorded into seven classes using the Recode utility of the GIS analysis.

For the supervised LCLU classification, 160 signatures were defined directly from the AOI of the Landsat image using the Signature Editor. A total of 25 signatures were created for each class except OTHER land, and only 10 signatures were created for OTHER land. After evaluating and merging the signatures, the supervised LCLU classification was performed using the maximum likelihood classifier (MLC) of the supervised classification. A classified image and a distance file were created. For the hybrid classification, there were four steps. The supervised classification was performed first. It generated a classified image and a distance file

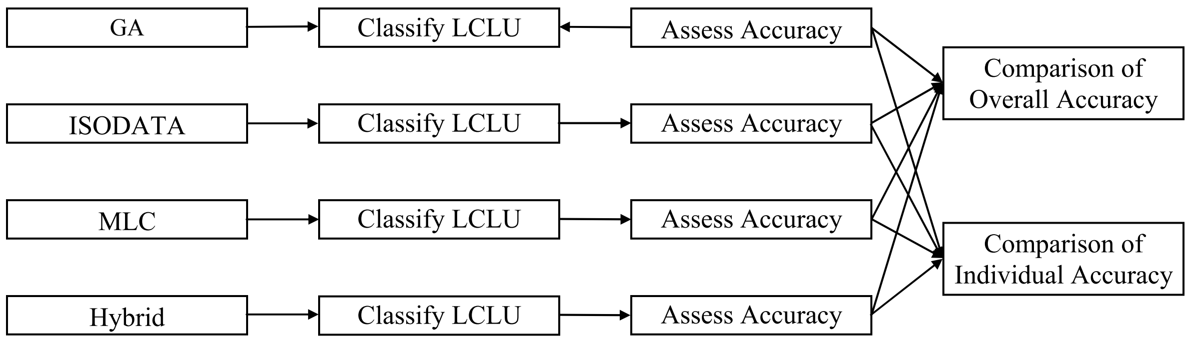


Figure 6.1. The workflow for comparing classification accuracy of different approaches (including GA, ISODATA, MLC, and hybrid of MLC and ISODATA)

using MLC and probabilities. “Threshold” was then run to create the unclassified classes using a 0.05 confidence level and process to file. These unclassified classes were selected, copied to raster AOI, and saved to an AOI file. Second, the signature file with 160 classes, output from the unsupervised classification, was checked to ensure that the standard deviation was less than 10% of the mean for Bands 3, 4, and 5, and that the data distributions were normal. Signatures failing these checks were eliminated from further analysis. The remaining signatures’ count was copied into “probability”. After normalization, the probability and above AOI were used for the supervised classification. The third step was to assign the signatures into different classes and recode them. The final step was to create a model to recode and combine the final supervised and unsupervised classification.

For GA-based classification, a set of training / testing data (20000 / 10000) was created from the resultant image of ISODATA. Since the resultant image has 160 classes and the smallest class has 313 pixels, the minimum amount of pixels was set to 100, and a stratified random sampling technique was selected. The pixels’ coordinates file was saved and used for taking training / testing samples from the AOI of Landsat image. After the random sampling, the data are converted into ARFF file that is supported by most data mining/machine learning software packages. In order to discover the spectral knowledge hidden in this data set, the ARFF file was input into GATree 2.0, a GA-based data mining / machine learning software. The resultant chromosome was represented as a decision tree, which was interpreted into classification rules. Those rules then were used to code an expert classifier, and perform the LCLU classification.

Finally, for each approach, an accuracy assessment was performed after finishing the LCLU classification. The way to assess classification accuracy has been described in the previous three chapters. The details of the procedure will not be repeated here.

6.2. Results

An image AOI (4,257,000 pixels) covering the entire Tickfaw watershed was created. It was classified into seven classes according to the classification scheme (Table 3.4). The major results include the classified images, error matrices, and accuracy assessments. As it has been described above, the AOI of Landsat image was classified using various approaches, including unsupervised, supervised, hybrid, and GA-based LCLU classification techniques. The accuracy assessment was based on the error matrix or confusion matrix. Error matrix is a square array of values, which cross-tabulates the number of sample spatial data units assigned to a particular category relative to the actual category as verified by the reference data. Conventionally, the rows of the error matrix represent the categories of the classification of the database, while the columns indicate the classification of the reference data (Lo and Yeung, 2007). There are several terms describing the accuracy: 1) omission error defines the error of exclusion; 2) commission error defines the error of misclassification; 3) producer accuracy is the probability of a sample being correctly classified, and it measures the error of omission; 4) user accuracy is the probability that a sample actually represents the real world, and it measures the error of commission; 5) overall accuracy represents the percentage of correctly classified data; 6) kappa value controls the overestimation / underestimation of overall accuracy and it tests the statistical significance of differences in different error matrices (Congalton, 1991; Lo and Yeung, 2007). The detailed results from every approach are provided as follows.

Figure 6.2 demonstrates the resultant images generated by different LCLU classification approaches. Based on visual estimation, URBAN and GRASS were overestimated, and FARM was underestimated using the ISODATA-based unsupervised LCLU classification. FOREST was overestimated, and URBAN was underestimated using MLC-based supervised LCLU classification. The resultant images derived from the hybrid and GA-based LCLU classification

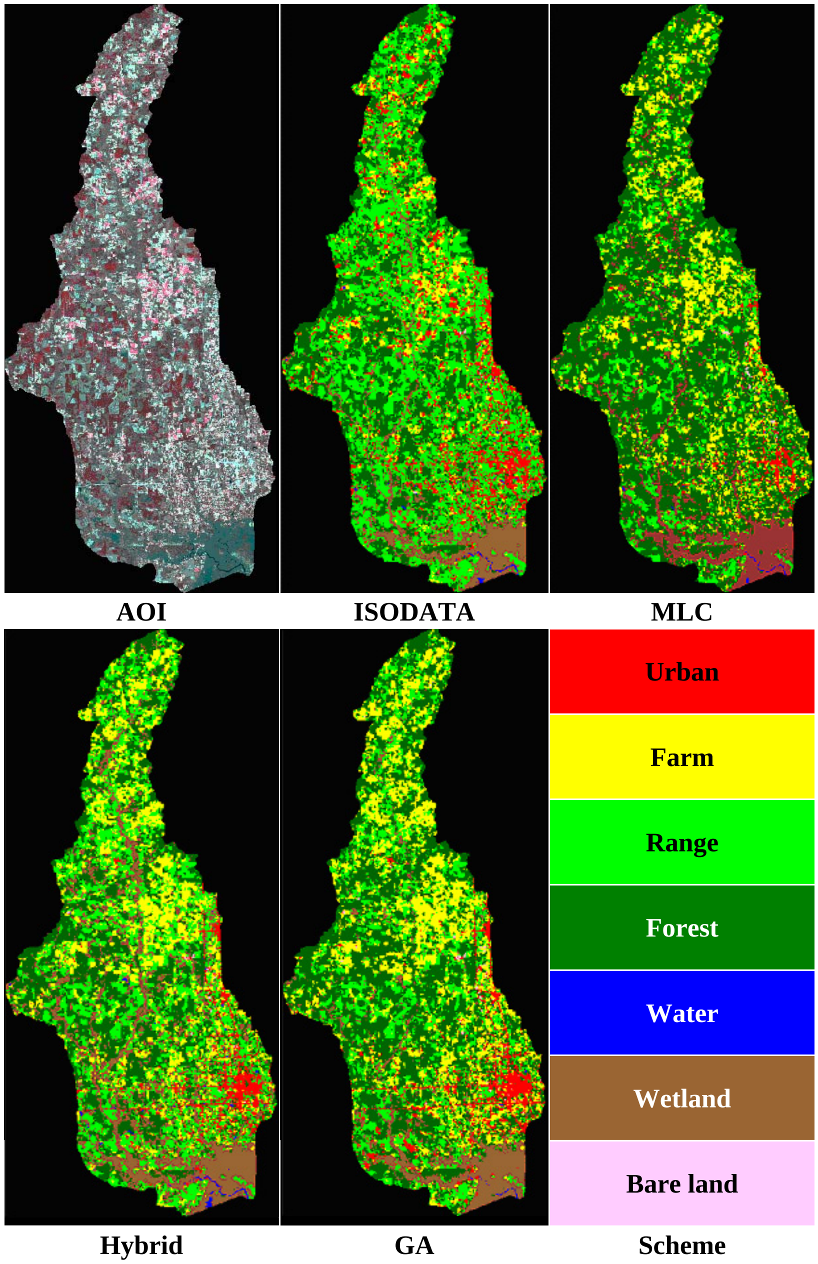


Figure 6.2. The results of land-cover and land-use classification using different algorithms-based approaches

approaches look similar. Those major overestimation and underestimation problems seem to have solved by both the hybrid and GA-based LCLU classification approaches. The detailed results can be found in the following sections.

6.2.1. ISODATA-based Unsupervised Land-Cover and Land-Use Classification

The first attempt made to classify LCLU was done using an ISODATA-based unsupervised classification. This approach does not require the user to specify any information about the land features contained in the images. The experiment was carried out in ERDAS Imagine. Compared to other approaches, this experiment provided the lowest overall accuracy (72.30%) and the lowest overall kappa value (0.6653) (Table 6.1). It also indicates that the classification of URBAN, GRASS, and OTHER did not perform well, because the agreement between the classified data and reference data is low. The kappa values of these three land features are only 0.4189, 0.3392, and 0.4893, respectively. The error matrix shows large errors of omission between URBAN and GRASS, FARM and URBAN, GRASS and URBAN, and OTHER and URBAN. The error matrix also shows large errors of commission between URBAN and FARM, GRASS and URBAN, GRASS and FARM, GRASS and FOREST, GRASS and WETLAND, and OTHER and URBAN. These results suggest that: 1) land feature URBAN and GRASS was overestimated considerably; 2) land feature FARM was underestimated considerably; and 3) land feature OTHER was considerably misclassified and omitted.

6.2.2. MLC-based Supervised Land-Cover and Land-Use Classification

The second attempt made to classify LCLU was done using MLC-based supervised classification. This approach requires the user to specify some information about the land features contained in the images. The experiment was carried out in ERDAS Imagine. Compared with other approaches, this experiment improved the overall accuracy (78.33%) and overall kappa value (0.7331) (Table 6.2), but the confusion among URBAN, GRASS, and FOREST still

Table 6.1. Error matrix and accuracy assessment of ISODATA-based approach

Sample Data	Reference Data							Row Total
	Urban	Farm	Grass	Forest	Water	Wetland	Other	
Urban	123	88	11	2	1	0	11	236
Farm	9	91	9	1	0	1	0	111
Grass	26	19	66	25	0	28	1	165
Forest	4	0	6	174	0	2	0	186
Water	0	0	0	1	63	6	0	70
Wetland	4	1	0	8	0	197	0	210
Other	9	0	0	0	0	0	9	18
Column Total	176	199	92	211	66	235	21	--

Class Name	Reference Total	Classified Total	Number Correct	Producer Accuracy	User Accuracy	Kappa Value
Urban	176	236	123	69.89%	52.12%	0.4189
Farm	199	111	91	45.73%	81.98%	0.7751
Grass	92	165	66	71.74%	40.00%	0.3392
Forest	211	186	174	82.46%	93.55%	0.9182
Water	66	70	63	95.45%	90.00%	0.8929
Wetland	235	210	197	83.83%	93.81%	0.9191
Other	21	18	9	42.86%	50.00%	0.4893
Total	1000	1000	723	--	--	--
Overall	--	--	--	**72.30%**		**0.6653**

remains. Table 6.2 also shows that the classification of GRASS and FOREST does not perform well. The Kappa values of these two land features are only 0.3502 and 0.5946, respectively. The error matrix shows large omission errors between URBAN and GRASS, URBAN and FOREST, and GRASS and FOREST. The error matrix also shows the high errors of commission between GRASS and URBAN, FOREST and GRASS, and FOREST and URBAN. These results indicate: 1) FOREST was considerably overestimated; 2) URBAN was considerably underestimated; and 3) GRASS was considerably misclassified and omitted.

6.2.3. Hybrid-Approach-based Land-Cover and Land-Use Classification

The third attempt made to classify LCLU is done using the hybrid classification technique, particularly the integration of ISODATA and MLC. The experiment is carried out in ERDAS Imagine. Compared with the previous two approaches, because this approach seeks to combine advantages from both to overcome the disadvantages from each other, it significantly improves the overall accuracy (81.90%) and overall kappa value (0.7791) (Table 6.3). Although the errors are reduced, they still exist. For example, the error among URBAN, FARM, and GRASS is apparent. The results also show that the classification of GRASS does not perform well. Its kappa value is only 0.4146. The error matrix shows that there are large omission errors between GRASS and FARM. The error matrix also shows large commission errors between GRASS and URBAN, GRASS and FARM, GRASS and FOREST, and GRASS and WETLAND. These results indicate that GRASS was still considerably misclassified and omitted.

6.2.4. GA-based Land-Cover and Land-Use Classification

The last attempt made to classify LCLU was done using an artificial intelligence-based classification technique, particularly the GA. Based on the given training / testing data, the approach seeks to find a global optimized solution by enlarging and searching the solutions' space. The experiment was carried out in ERDAS Imagine and GATree. Compared with other

Table 6.2. Error matrix and accuracy assessment of MLC-based approach

Sample Data	**Reference Data**							**Row Total**
	Urban	**Farm**	**Grass**	**Forest**	**Water**	**Wetland**	**Other**	
Urban	96	0	1	1	0	0	1	99
Farm	14	177	13	0	0	0	4	208
Grass	36	12	41	3	1	7	0	100
Forest	26	10	36	185	0	15	0	272
Water	0	0	0	0	58	2	0	60
Wetland	4	0	1	22	7	210	0	244
Other	0	0	0	0	0	0	16	16
Column Total	176	199	92	211	66	235	21	--

Class Name	**Reference Total**	**Classified Total**	**Number Correct**	**Producer Accuracy**	**User Accuracy**	**Kappa Value**
Urban	176	99	96	54.55%	96.97%	0.9632
Farm	199	208	177	88.94%	85.10%	0.8139
Grass	92	100	41	44.57%	41.00%	0.3502
Forest	211	272	185	87.68%	68.01%	0.5946
Water	66	60	58	87.88%	96.67%	0.9643
Wetland	235	244	210	89.36%	86.07%	0.8179
Other	21	16	16	76.19%	100.00%	1.0000
Total	1000	1000	783	--	--	--
Overall	--	--	--	**78.30%**		**0.7331**

Table 6.3. Error matrix and accuracy assessment of hybrid-based approach

Sample Data	Reference Data							Row Total
	Urban	Farm	Grass	Forest	Water	Wetland	Other	
Urban	137	4	5	2	0	0	4	152
Farm	11	179	22	1	0	2	0	215
Grass	16	16	52	13	0	14	0	108
Forest	6	0	11	162	0	1	0	180
Water	0	0	0	1	64	9	0	74
Wetland	4	0	2	32	0	208	0	246
Other	1	0	0	0	0	0	17	18
Column Total	176	199	92	211	66	235	21	--

Class Name	Reference Total	Classified Total	Number Correct	Producer Accuracy	User Accuracy	Kappa Value
Urban	176	152	137	77.84%	90.13%	0.8802
Farm	199	215	179	89.95%	83.26%	0.7910
Grass	92	111	52	56.52%	46.85%	0.4146
Forest	211	180	162	76.78%	90.00%	0.8733
Water	66	74	64	96.97%	86.49%	0.8553
Wetland	235	246	208	88.51%	84.55%	0.7981
Other	21	18	17	80.95%	94.44%	0.9433
Total	1000	1000	819	--	--	--
Overall	--	--	--	**81.90%**		**0.7791**

approaches, this experiment improved the overall accuracy significantly (84.70%) and the overall kappa value was 0.8119 (Table 6.4). Although the errors were reduced, they still exist. For example, the confusions among FARM, GRASS, and FOREST are apparent. The results also show that the classification of GRASS is still less significant. Its kappa value is improved, but it was only 0.5010. The error matrix shows large omission errors between GRASS and FARM. However, the error matrix shows no considerable errors of commission. These results indicate that GRASS was still considerably omitted.

6.3. Discussion

The comparison study can be carried out differently depending on the objective of the study. They can be based on: 1) advantages and disadvantages (*e.g.*, data preprocessing, training data sampling, and algorithms), 2) ease of operation, and 3) overall and individual level of classification accuracy. Because one of the objectives of this study is to compare the traditional and GA-based LCLU classification approaches, the results were discussed in terms of all these three aspects.

6.3.1. Advantages and Disadvantages

The classification approaches used in this study can be categorized into statistical or non-statistical, and parametric and non-parametric (Table 6.5). ISODATA is a non-parametric statistical approach. MLC is a parametric statistical approach. Since the hybrid approach combines both ISODATA and MLC, it is also statistics. GA is a heuristic approach based on artificial intelligence. Different approaches have different assumptions, advantages, and disadvantages. When the data distribution is normal, MLC performs well in LCLU classification. When the data distribution is not normal, the GA-based approach performs better than the other approaches in terms of LCLU classification. ISODATA requires subsequent identification of land features. Although the hybrid approach seeks to integrate the advantages from both

Table 6.4. Error matrix and accuracy assessment of GA-based approach

Sample Data	**Reference Data**							**Row Total**
	Urban	**Farm**	**Grass**	**Forest**	**Water**	**Wetland**	**Other**	
Urban	142	3	10	3	0	1	2	161
Farm	17	184	29	3	0	1	0	234
Grass	3	7	35	8	0	11	0	64
Forest	6	2	16	196	0	8	0	228
Water	0	0	0	0	58	0	0	58
Wetland	2	0	2	1	8	213	0	226
Other	6	3	0	0	0	0	19	28
Column Total	176	199	92	211	66	235	21	--

Class Name	**Reference Total**	**Classified Total**	**Number Correct**	**Producer Accuracy**	**User Accuracy**	**Kappa Value**
Urban	176	161	142	80.68%	88.20%	0.8568
Farm	199	234	184	92.46%	78.63%	0.7332
Grass	92	64	35	38.04%	54.69%	0.5010
Forest	211	228	196	92.89%	85.96%	0.8221
Water	66	58	58	87.88%	100.00%	1.0000
Wetland	235	226	213	90.64%	94.25%	0.9248
Other	21	28	19	90.48%	67.86%	0.6717
Total	1000	1000	847	--	--	--
Overall	--	--	--	**84.70%**		**0.8119**

Table 6.5. Characteristics of different land-cover and land-use classification approaches

Algorithms	Algorithms Types	Assumptions	Advantages	Disadvantages
ISODATA	Statistical, non-parametric	Data distribution is normal	No training data required, automatically clustering	Require subsequent identification of land features, no user control, low accuracy, assumption may not be valid
MLC	Statistical, parametric	Data distribution is normal	User has control on procedure, good accuracy	Training data important, the assumption may not be valid
HYBRID	Statistical, both non-parametric and parametric	Data distribution is normal	Combine the advantage from both ISODATA and MLC, very good accuracy	Training data important, the assumption may not be valid , and cost time
GA	Non-statistical, non-parametric, artificial intelligence,	Data distribution is not important	Random search and global optimization, very good accuracy	Require good understanding of user, large training data, and large training time

ISODATA and MLC (Richards and Jia, 1993), the disadvantages from both may still affect the hybrid's performance. If the data distributions are not normal, the recommended selection order from the first to the last is GA > Hybrid > MLC > ISODATA.

6.3.2. Ease of Operation

The different classification approaches had different requirements on the training data, required different classification procedures, and consumed different amounts of time (Table 6.6). ISODATA did not need training and required less time. GA needed a large amount of training data, went through a complex classification procedure, and required a long time. When an application must be done within very limited amount of time, ISODATA is the most efficient approach to be selected. When time is not a problem, GA-based approach should be selected. For any application, if time is not a problem, the recommended selection order from the first to the last is GA > Hybrid > MLC > ISODATA.

6.3.3. Overall Classification Accuracy

The comparison of the classification accuracy at the overall level was based on the overall accuracy and overall kappa statistics. The different classification approaches performed differently in LCLU classification (Figure 6.1). The order of their performance from high to low was GA > Hybrid > MLC > ISODATA, the GA-based approach had the best performance with the overall LCLU classification accuracy of 84.70% and the kappa value of 0.8119. The ISODATA approach had the worst performance with the overall LCLU classification accuracy of 72.30% and a kappa value of 0.6653. The other two approaches have moderate performance. This is not surprising, because GA is an approach specially designed for randomly searching and globally optimizing the solution. It has no assumption on data characteristics. The other approaches are statistics-based, and all of them assume that the distribution of data is normal. Unfortunately, this is not always valid.

Table 6.6. The operations of different land-cover and land-use classification approaches

Algorithms	Training Data	Operation	Time
ISODATA	No training data required, but 160 clusters were created and labeled.	Specify number of clusters, confidence level, label clusters into classes.	Less than 4 Hours
MLC	160 signatures were taken and trained using AOI	Create, evaluate, and merge signatures, use them for classification	About 6 Hours
HYBRID	160 clusters and 160 signatures were taken using AOI tool	Perform ISODATA and MLC based classification respectively, combine the results from both.	About 8 hours
GA	Create 20000 training samples / 10000 testing samples using stratified random sample and the 160 clusters image from ISODATA.	Take training samples, discover knowledge, code expert classifier, classify image.	At least 48 hours

6.3.4. Individual Classification Accuracy

The comparison of classification accuracy at the class level was based on producer accuracy, user accuracy, and kappa value for each class. The different classification approaches performed differently for different LCLU features (Figures 6.3 through 6.10). Based on these figures, a performance evaluation is made, and the order of their performance on each land feature, from high to low, is illustrated (Table 6.10). The possible reasons for these results are argued as follows.

For the land feature URBAN, the GA-based approach did not perform better than supervised and hybrid approaches because the complexity in urban areas is high, and it usually has various land features. The way to sample training data in MLC, hybrid, and GA-based approaches is different. In the GA-based approach, because training samples were taken from different bands respectively, a coordinate file was needed. Therefore, the pixel-by-pixel-based sampling approach was used. But in the MLC and hybrid approaches, there were at least two ways to take the training data: area-by-area and pixel-by-pixel. For this dissertation research, in order to save time, the training data of MLC and Hybrid were automatically taken area-by-area. Obviously, the training data in MLC and Hybrid approaches are more homogeneous.

For the land feature FARM, the GA-based approach performed had the highest producer accuracy, but the user accuracy is the worst because FARM was usually mixed with URBAN and GRASS, so it was difficult to distinguish them visually. When a pixel-by-pixel approach was used to take training data for GA-based LCLU classification, it is hard to ensure 90% accuracy. On the other hand, an area-by-area approach was automatically performed to take training data for MLC-based and hybrid-based classification, and these training data are more homogeneous.

For the land feature GRASS, although the GA-based approach performed better than other approaches used in this study, the accuracy was low: producer accuracy of 38.04%, user

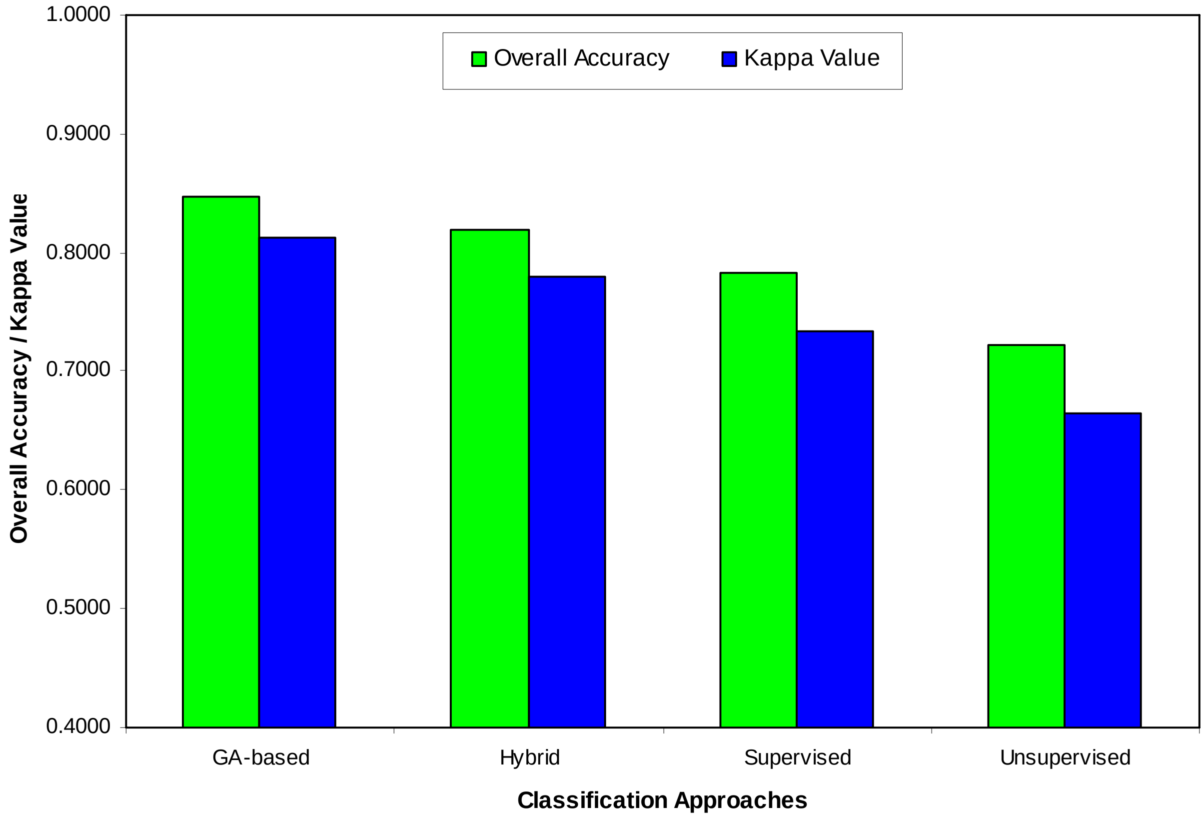


Figure 6.3. Comparison of the overall accuracy and kappa value of different classification approaches

accuracy of 54.69%, and kappa value of 0.5010 because GRASS was mixed with most features, such as URBAN, FARM, FOREST, and WETLAND. All approaches used in this research had a difficult time classifying it correctly. Although the GA-based approach is promising, it requires much time. In this experiment, the GA-based model involved 2000 generations with 1000 individuals, and it required about three days to finish.

For the land feature FOREST, the GA-based approach did perform better than ISODATA and hybrid, with producer accuracy of 92.89%, user accuracy of 85.06%, and kappa value of 0.8221. FOREST was the largest feature in the study area. When various land features are mixed together, to distinguish FOREST is usually easier than to distinguish GRASS.

Although the area-based sampling approach still had an advantage over the pixel-based sampling approach, the difference was reduced because the FOREST feature was easy to identify. In this case, the training data from both sampling approaches were highly homogeneous. For the land feature WATER and WETLAND, the reason will be similar.

For the land feature OTHER, the GA-based approach did not perform better than MLC and hybrid. This may be due to the fact that OTHER was small and usually mixed with URBAN, so it was difficult to distinguish them visually. When a pixel-by-pixel approach was used to collect training data for GA-based LCLU classification, it was difficult to ensure 90% accuracy rate. On the other hand, an area-by-area approach was performed automatically to collect training data for the MLC-based and hybrid-based classification, and these training data are obviously highly homogeneous.

In general, knowledge about the study area is very important for all of approaches. For ISODATA, and without being familiar with study area, it will be difficult to label the clusters into the classes. For the other three approaches, without enough knowledge about the study area, the quality of training data will be low. In addition to the knowledge about the study area,

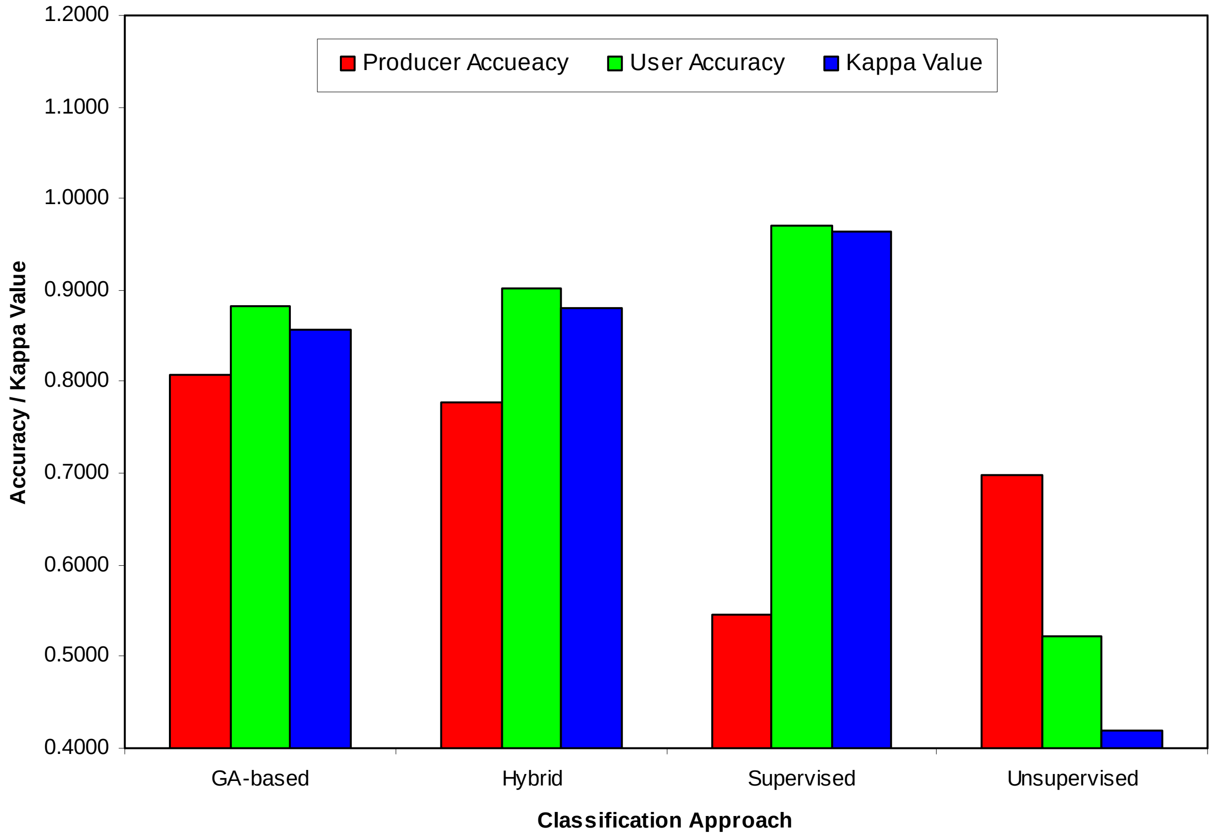


Figure 6.4. Comparison of the URBAN accuracy and kappa value of different classification approaches

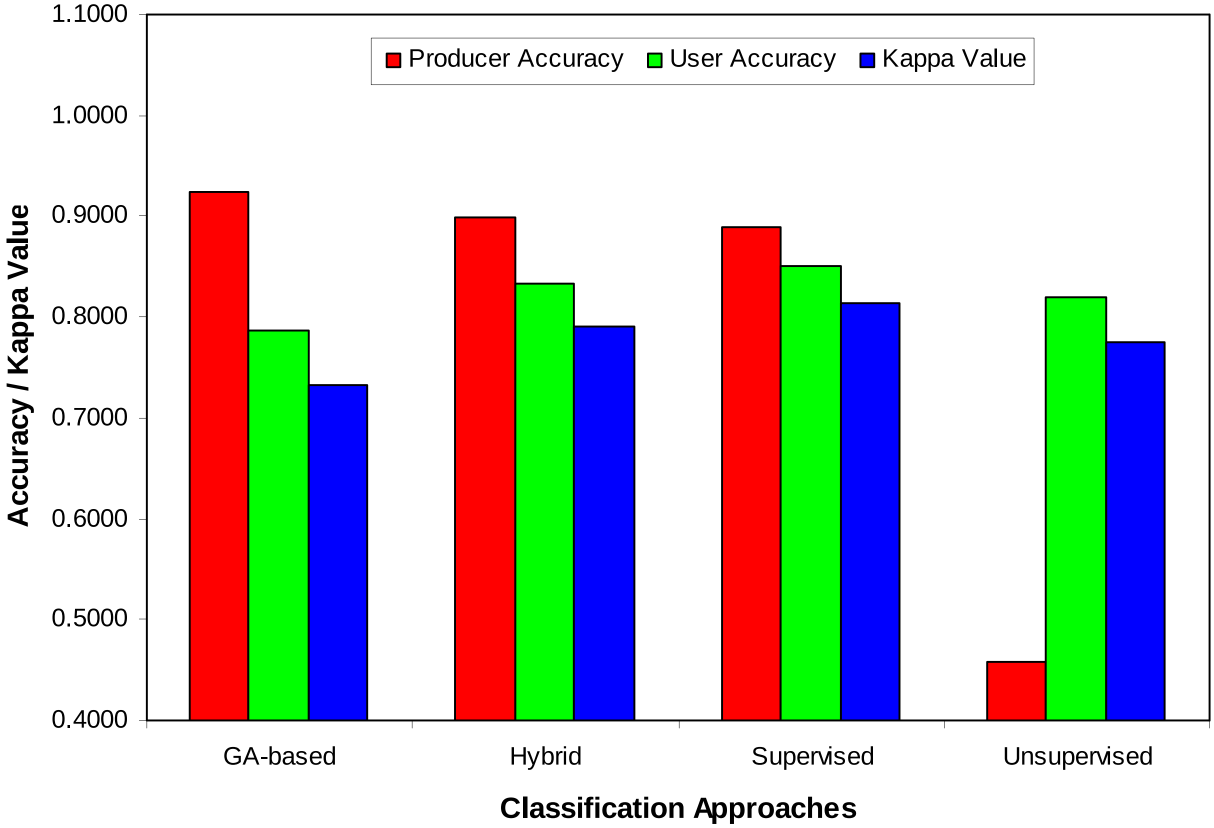


Figure 6.5. Comparison of the FARM accuracy and kappa value of different classification approaches

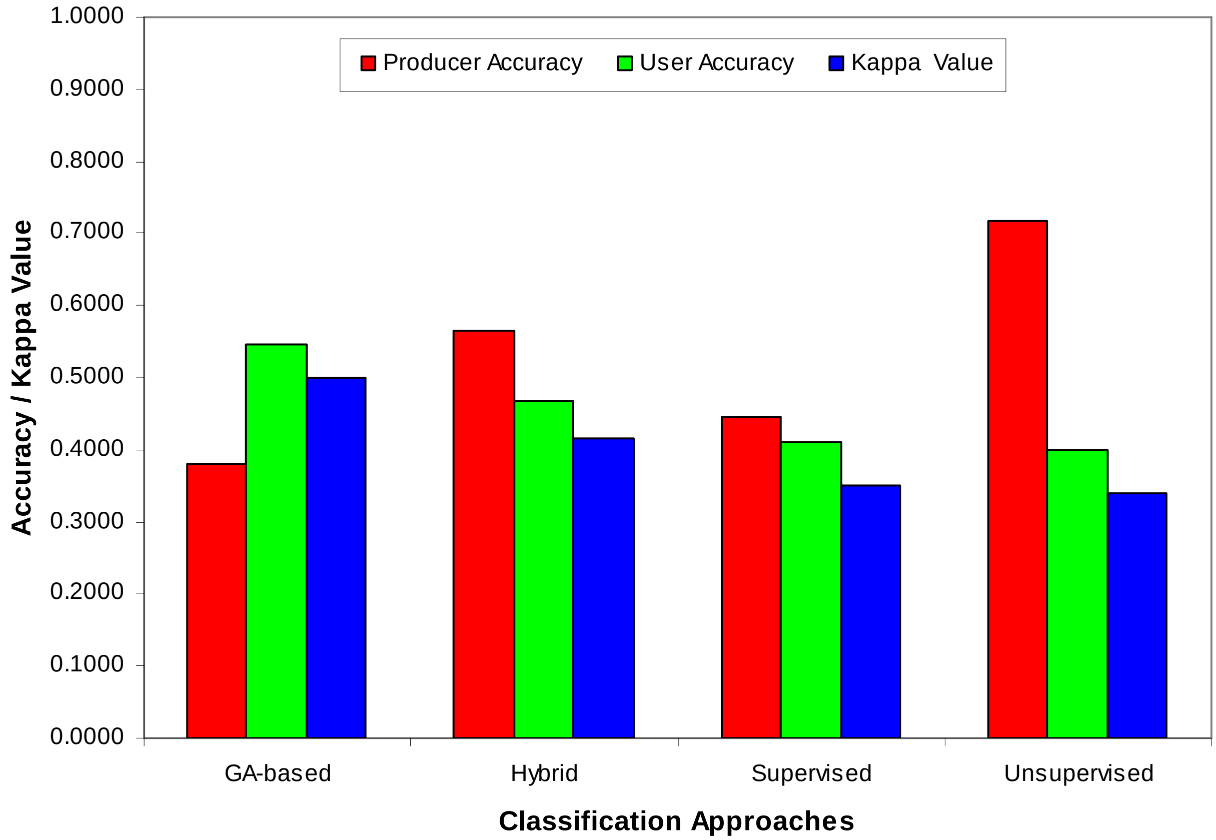


Figure 6.6. Comparison of the GRASS accuracy and kappa value of different classification approaches

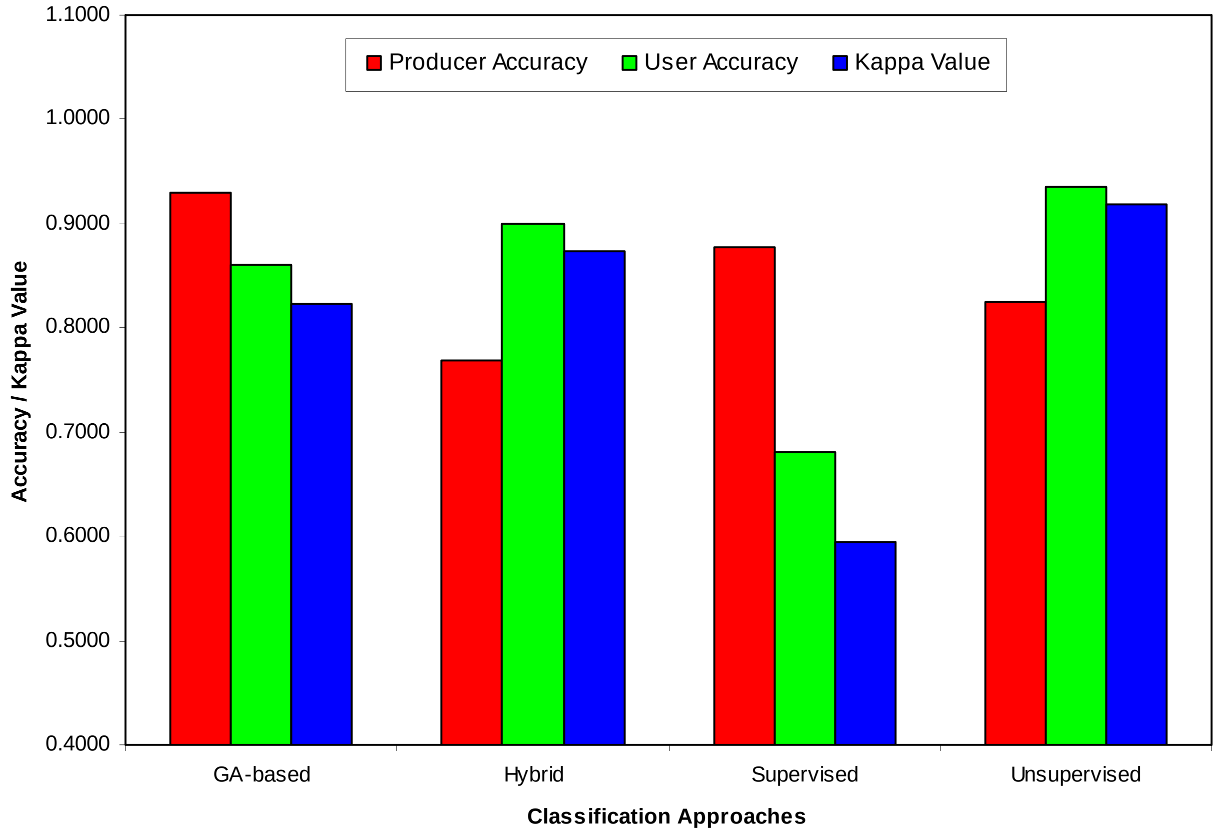


Figure 6.7. Comparison of the FOREST accuracy and kappa value of different classification approaches

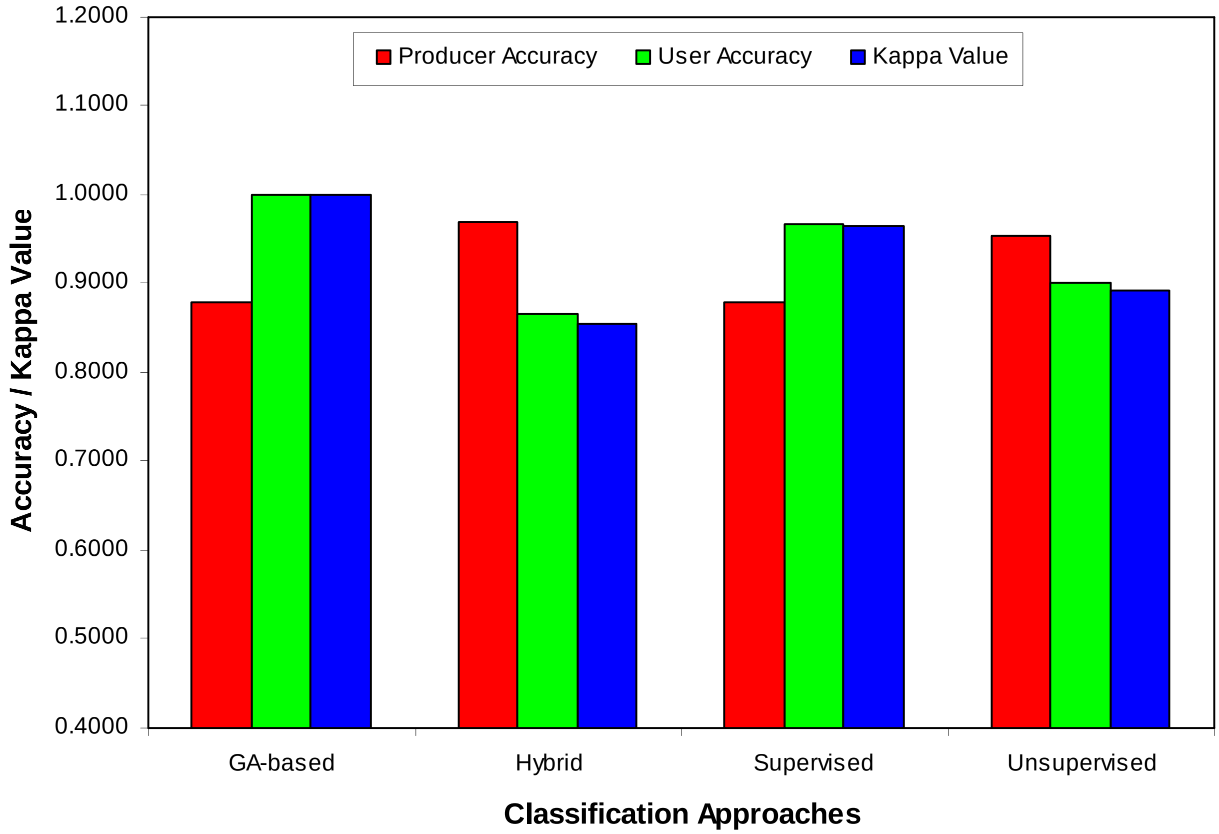


Figure 6.8. Comparison of the WATER accuracy and kappa value of different classification approaches

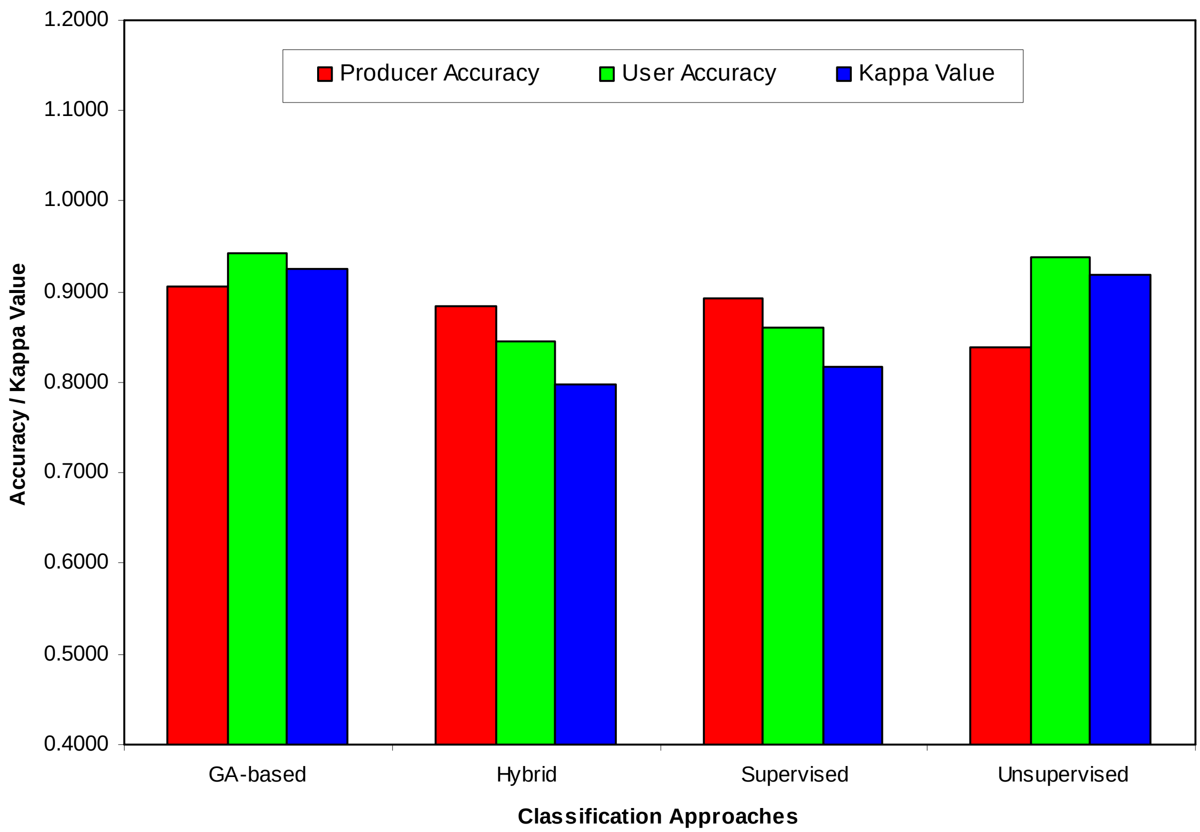


Figure 6.9. Comparison of the WETLAND accuracy and kappa value of different classification approaches

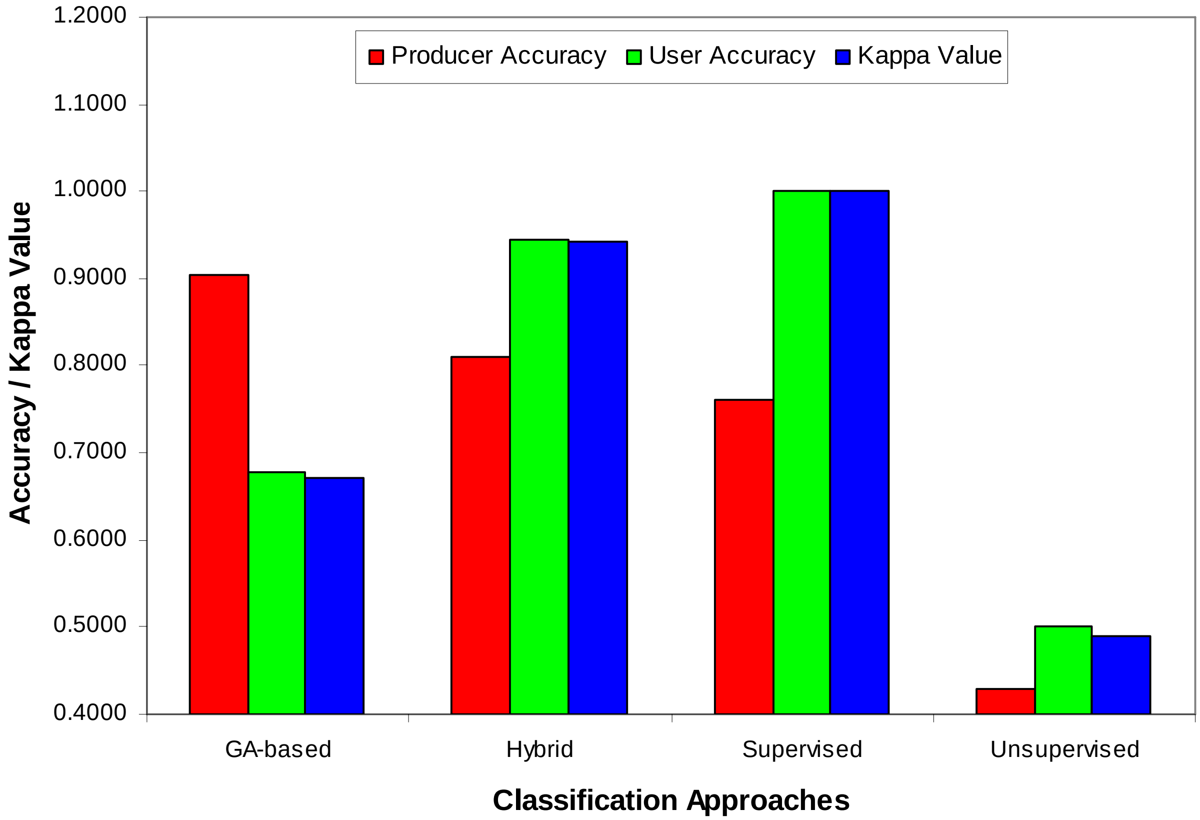


Figure 6.10. Comparison of the OTHER accuracy and kappa value of different classification approaches

Table 6.7 The performance order of different approaches
on each class (from high to low)*

Class Name	Producer Accuracy	User Accuracy	Kappa Value
URBAN	G, H, U, S	S, H, G, U	S, H, G, U
FARM	G, H, S, U	S, H, U, G	S, H, U, G
GRASS	U, H, S, G	G, H, S, U	G, H, S, U
FOREST	G, S, U, H	U, H, G, S	U, H, G, S
WATER	H, U, G, S	G, S, U, H	G, S, U, H
WETLAND	G, S, H, U	G, U, S, H	G, U, S, H
OTHER	G, H, S, U	S, H, G, U	S, H, G, U
*G – GA-based Classification; H – hybrid classification; S – supervised classification; and U – unsupervised classification.			

another reason that the performance of ISODATA, MLC, and hybrid approach with spectral data approach was not effective may be due to the distribution of data. These three approaches assume that the data distribution for each class is Gaussian (normally distributed). However, this assumption may not be always valid (Hutchinson, 1982). Sometimes the image data have bi- or multi-modal distribution (Huang and Jensen, 1997). Moreover, the reason that the performance of the GA-based approach with spectral data approach was not always effective for each class may be that GA is a random search and global optimization technique. It does not look for local optimization (*e.g.*, good for particular land feature classification for a particular area); rather, it seeks global optimization (*e.g.*, good for all land features classification for entire study area).

6.4. Hypothesis # 3 Review

The third research question as presented in Chapter 1 was whether GA-based LCLU classification is more accurate than maximum likelihood classifier (MLC) based supervised classification, iterative self-organizing data analysis technique (ISODATA) based unsupervised classification, or the hybrid of both MLC and ISODATA. The third research hypothesis presented in Chapter 1 was stated as follows:

Ha: GAs-based LCLU classification is more accurate than MLC-based supervised classification, ISODATA-based unsupervised classification, and the hybrid of MLC and ISODATA, respectively.

Based on the experiment performed in this chapter, the results show that there were some differences among the unsupervised, supervised, hybrid, and the GA-based LCLU classification. With the consideration of the characteristics of those approaches, GA-based approach is an artificial intelligence and non-parametric approach, and it has no assumption on the distribution of data. Those traditional approaches are statistics-based approaches, and they have assumption on the distribution of data (normal distribution). This assumption will limit the application of traditional approaches. With the consideration of the operation of those approaches, ISODATA needs no training data and less time. Other approaches require training data, and the GA-based approach requires more time. With the consideration of the accuracy of those approaches, at the overall level, the GA-based approach has the highest overall accuracy and kappa value; at the individual level, the GA-based approach has the highest producer accuracy, user accuracy, and Kappa value in most land features classification. Therefore, it can be concluded that the GA–based is more accurate than traditional approaches.

6.5. Summary

As it has been stated, the primary goal of this study is to compare the traditional and GA-based LCLU classification approaches in terms of advantages and disadvantages, ease of operation, overall and individual levels of classification accuracy. The characteristics of different approaches were stated in terms of algorithm types, assumptions, advantages, and disadvantages. The operation of those approaches was compared in terms of training data, classification procedure, and time consumption. Finally, the performance of those approaches was evaluated in terms of the overall and individual-level accuracy. In summary,

- The characteristic of those algorithms – the GA-based approach is an artificial intelligence and non-parametric approach, and it has no assumption on the distribution of data. Those traditional approaches are statistics-based approaches, and they have assumption on the distribution of data (normal distribution). This assumption will limit the application of traditional approaches.
- The operation of those algorithms – ISODATA is an approach which needs no training data and less time. Other approaches require training data, and the GA-based approach requires a lot of time.
- The performance of those algorithms – At the overall level, GA-based approach has

the highest overall accuracy and kappa value; At the individual level, the GA-based approach has the highest producer accuracy, user accuracy, and kappa value in most land feature classifications.

- Finally, it can be concluded that the GA-based is more accurate than the traditional approaches.

CHAPTER 7 PROCESS-ORIENTED LAND-COVER AND LAND-USE CHANGE DETECTION USING GENETIC PETRI NET

7.1. Experiment #4 Design

If LCLU classification can be recognized as the procedure of turning pixels into patterns, change detection can be recognized as the procedure of turning patterns into processes. In the previous three chapters, the major issue examined was about GA-based LCLU classification, and the goal was to find the spatial patterns of LCLU. The major issue in this chapter is about genetic Petri net (PN) based change detection, and the goal is to discover the process models of LCLU change. The purposes of discovering process model are to explain how and why LCLU change, and to provide transition rules for cellular automata (CA) based spatial modeling.

The experiment was designed to examine the relationships between genetic parameters and process-oriented LCLU change detection, especially to discover a better GA-parameter configuration for identifying change process models (Figure 7.1). Six images were acquired during 1990-2002 (Figure 3.1). These images were classified using GA-based LCLU classification techniques described in Chapter 4 and 5. The optimized configuration of genetic parameters and image parameters is: number of generation 5000, population size 1000, crossover rate 99.00%, mutation rate 0.50%, generation gap 50.00%, spatial resolution 30 m, training / testing data size 20000 / 10000, and the data included Landsat TM/ETM image (Bands 1-5, and Band 7) and DEM. Table 7.1 demonstrates the LCLU and accuracy assessment, and Figure 7.2 shows each layer of LCLU.

After classification, a LCLU stack was created. In order to mine the processes of LCLU change, an event log file should be compiled first. An event represents a pixel holding a value at a particular time and a particular location. The value held by a particular pixel may change over time. A new event occurs when the pixel's value changes. A set of time serial events constitutes a process instance, and a set of process instances constitutes a log. In this experiment, 7992

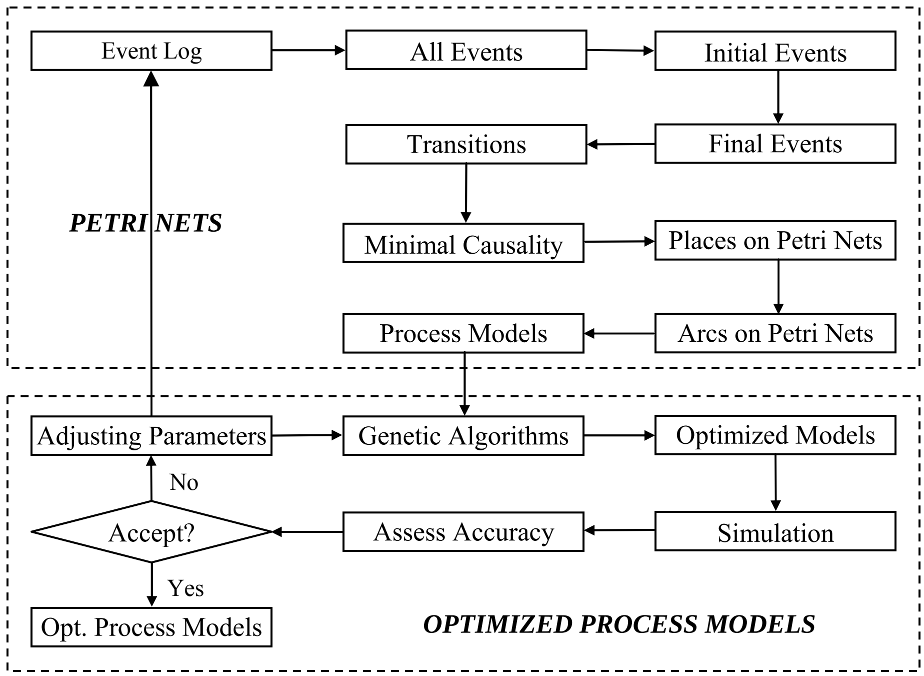


Figure 7.1. The workflow for optimizing genetic parameters
in genetic Petri nets-based process mining

Table 7.1. Land-cover and land-use change in the Tickfaw River watershed

Class Name	**1990**	**1992**	**1995**	**1997**	**2000**	**2002**
URBAN	49262	57816	63629	83887	106371	134114
FARM	212224	260249	291328	274723	261460	261405
GRASS	305354	377267	446783	477951	492553	473060
FOREST	843921	717630	627559	594992	573805	566731
WATER	4818	4750	4676	4830	4801	4725
WETLAND	688170	688106	674603	671708	668382	667511
OTHER	6875	4806	2046	2530	3246	3072
Unit: pixel Spatial resolution: 30 meters						

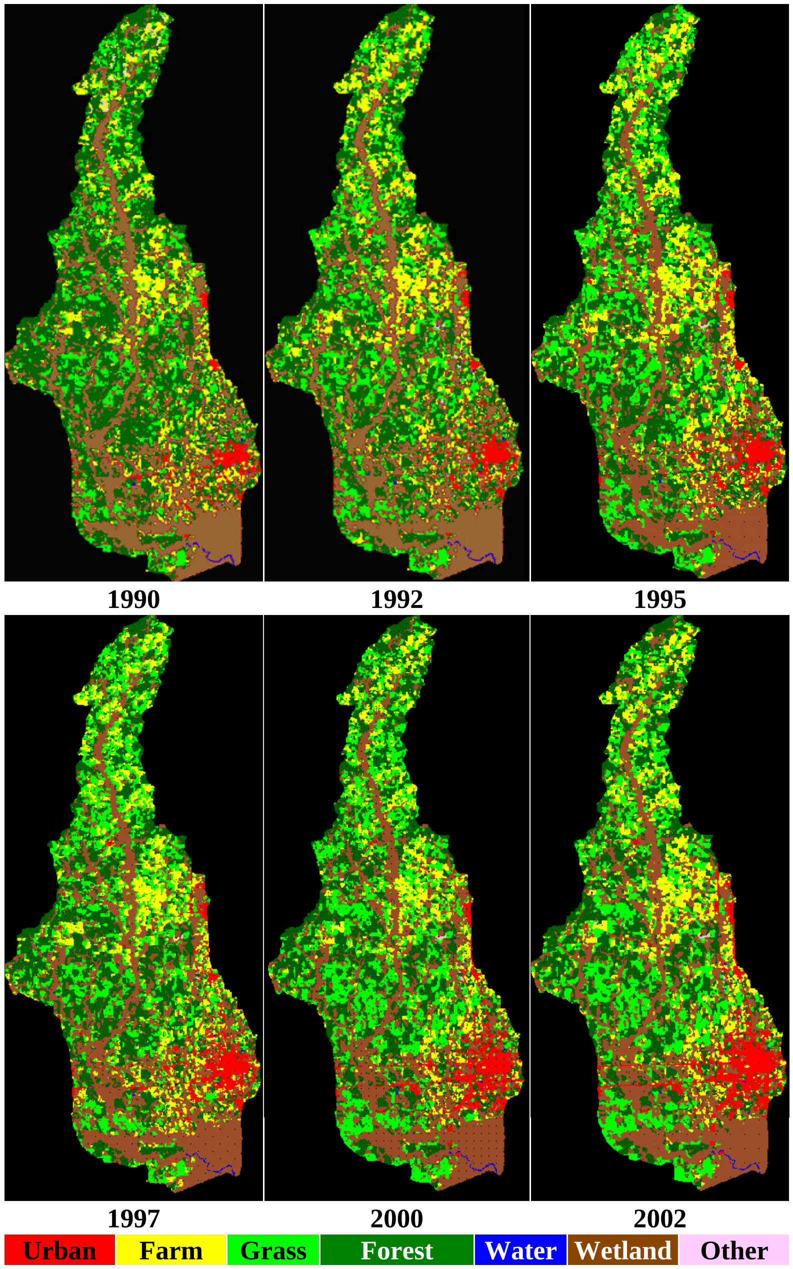


Figure 7.2. Land-cover and land-use classified images of Tickfaw watershed from 1990 to 2002

pixels were randomly taken from each layer of the stack, and each pixel had the same coordinate on different layers. Based on these samples, an XML database was created and was used as an input file for the process mining. In this study, the process mining was carried out on the environment of ProM 3.5 and ERDAS Imagine 9.1.

There were 39 tests designed for this experiment. The same XML database was used in every test. Those tests were categorized into 6 groups. There were 6 tests for each of the first 5 groups (Table 7.2), and the last group has 9 tests. Different group tests were used for different purposes. The first group was used to test the impact of the number of generations; The second group was used to test the impact of population size; The third group was used to test the impact of crossover rate; The fourth group was used to test the impact of mutation rate; The fifth group was used to test the impact of elitism rate; and the last group was used to discover urban growth process models. The default configuration is: generation 200, population 100, crossover 0.79, mutation 0.2, and elitism 0.02. When testing one parameter, the rest of them remained at default values. The optimized configuration was developed after finishing the first 5 groups of tests.

Since process mining is a very new research area, there is no standard method to examine the accuracy of the discovered processes. In this study, the accuracy was assessed by using average fitness, best fitness, and process instances coverage. If a process model was developed from a population with a high average fitness, its fittest was high, and it covered most process instances, then this process model was recognized as a good one.

The assumption was that the process model with the high average fitness, the high fittest, and the high instances coverage can correctly represent more process instances than a process model with low average fitness, low fittest, and low instances coverage. Since the study dealt with 6 years time series images and 7 LCLU features, theoretically, there were 7^6 (or 117649) process models that can be defined. Those models were categorized into 7^2 (or 49) groups based on the start state and end state (Figure 7.3). For example, the first group of processes theoretically includes 1 to 1, 1 to 2, 1 to 3, 1 to 4, 1 to 5, 1 to 6, and 1 to 7. As it has been described in Table 3.4, these digital numbers represent the particular LCLU features, for example, 1 represents URBAN, and 7 represents OTHER. Although some of them, for example, 1 to 2, and 1 to 5, can be eliminated based on the reality, most of them cannot be defined easily when considering the intermediate states.

The many-to-many relationships among the LCLU features during 1990-2002 were discovered using the log file in XML format (Figure 7.4). This process model was based on only 1000 samples and it was not pruned yet. Usually, the pruning procedure is used to simplify the process model. Although the model was derived from only 1000 samples, the great complexity of LCLU change process can be demonstrated. The digital number in each box represents the number of pixels held by a LCLU feature, and the digital number near to each arc represents the number of pixels changed from one LCLU feature to another LCLU feature. Due to the problem complexity and the resource limitation, the accuracy assessment was performed on the 1000 pixel samples for all LCLU features.

After figuring out the optimized configuration of genetic parameters, 7 of the 49 groups process models were created, pruned, and interpreted for analyzing urban growth. Theoretically, the LCLU changes related to urban growth include URBAN to URBAN, FARM to URBAN, GRASS to URBAN, FOREST to URBAN, WATER to URBAN, WETLAND to URBAN and OTHER to URBAN. An urbanized area may be forested or grassed after many years. Therefore, the URBAN to FOREST and URBAN to GRASS will be two additional LCLU change process related to the urban growth. These nine groups of LCLU change process models were evaluated in terms of average fitness, best fitness, and the coverage of process instances. Finally, these nine groups of LCLU change process models were implemented for representing urban growth.

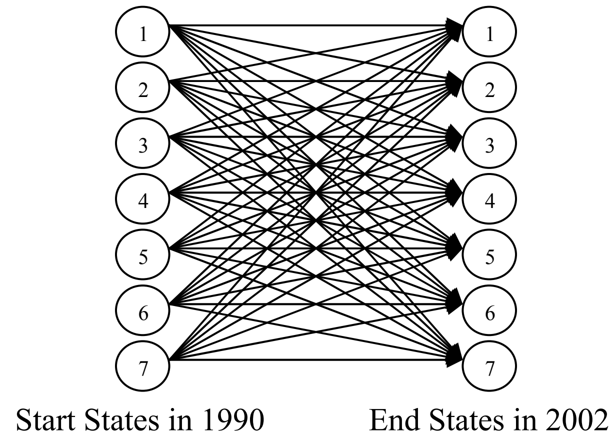


Figure 7.3. Theoretical group structure of change processes

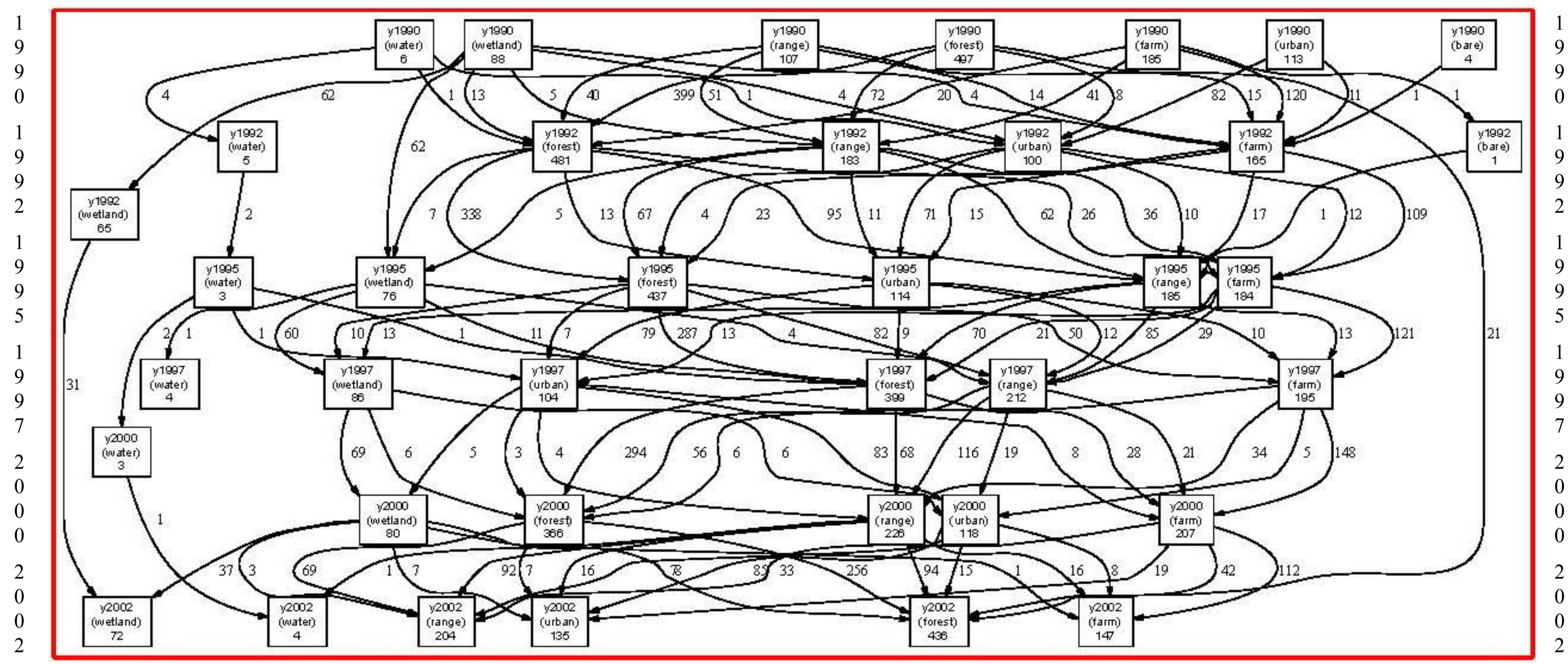


Figure 7.4. The complexity of land-cover and land-use change (the number in the box represents how many pixel for a particular land feature; the number on the arrow represents how many pixel transferred from one land-cover and land-use feature to another)

7.2. Results

This experiment focuses only on a number of problems in the GA-PN-based process mining. The problems include premature convergence, crowding, local optimization, and instability. These problems are usually caused by improper setting of GA parameters, such as the number of generations, population size, crossover rate, mutation rate, and elitism rate. When these problems occur, the performance of GA-PN-based process mining will not be good, and the resultant process models cannot discover the model which correctly represents most process instances.

Three measures were used to assess the accuracy of process mining: 1) the AVERAGE FITNESS of population is the sum of individuals' fitness divided by the total number of process models; 2) the FITTEST (or BEST FITNESS) defines an individual with the highest fitness, and it is the number of properly parsed process instances divided by the total number of process instances in a given event log; and 3) INSTANCES COVERAGE refers to the amount of process instancees that can be represented by the process model. The major results are given in Table 7.2. The goal of the first 5 groups of tests is to examine the impact of GA-parameters on process mining and to find the effective configuration of GA parameters for the process mining. After generating a set of process models, the best process model was selected and pruned at the rate of 1% - 10%. Pruning is a procedure to generalize the process model (or Petri net) and to eliminate the possible unreasonable processes. There is no standard method to determine the pruning rate. In this study, two criteria of setting up pruning rate were: 1) minimizing the number of unreasonable processes in the final models, 2) keeping the reasonable processes as many as possible. The process models discovered under different configurations were evaluated using average fitness, best fitness, and process instance coverage. Process models with high fitness and high coverage are expected to perform well.

Table 7.2. Relationships between genetic parameters and process mining

	Case ID	Parameters Setting*	Average Fitness	Best Fitness	Instances Coverage***
Generation Size	G1	50	0.3195	0.8355	46
	G2	100	0.4696	0.8720	96
	G3	200	0.3537	0.9016	97
	G4	300	0.3279	0.9036	132
	G5	400	0.3426	0.9048	116
	G6	**500**	**0.3491**	**0.9118**	**133**
Population Size	P1	50	0.4358	0.8952	70
	P2	100	0.3537	0.9016	97
	P3	150	0.3566	0.9015	71
	P4	200	0.4539	0.9113	67
	P5	250	0.5937	0.9281	83
	P6	**300**	**0.6649**	**0.9337**	**101**
Crossover Rate	C1	0.09	0.3920	0.8768	57
	C2	0.19	0.4544	0.8768	65
	C3	0.39	0.3224	0.9052	52
	C4	**0.59**	**0.5840**	**0.9321**	**108**
	C5	0.79	0.3537	0.9016	97
	C6	0.99	0.6110	0.9270	85
Mutation Rate	M1	0.005	0.9642	0.9666	221
	M2	0.01	0.9547	0.9723	204
	M3	**0.05**	**0.9551**	**0.9727**	**208**
	M4	0.10	0.8793	0.9542	194
	M5	0.20	0.3537	0.9016	97
	M6	0.30	0.1664	0.8552	1
Elitism Rate	E1	0.01	0.5043	0.9054	70
	E2	0.02	0.3537	0.9016	97
	E3	**0.04**	**0.6460**	**0.9251**	**135**
	E4	0.10	0.5866	0.9127	98
	E5	0.20	0.6913	0.9148	71
	E6	0.40	0.6910	0.9267	78
Recommendation**		Optimization	0.9511	0.9810	207

*Default configuration included: generation 200, population 100, crossover 0.79, mutation 0.2, elitism rate 0.02, and pruning rate 2%.
**Recommendation configuration included: generation 500, population 300, crossover 0.59, mutation 0.05, elitism rate 0.04, and pruning rate 2%.
***There were totally 218 process instances.

7.2.1. Genetic Parameters

The first group of tests in this experiment was designed for examining the impact of the number of generations on process mining. The results indicate that the value of best fitness and process instance model coverage increases as the number of generations increases (Table 7.2). Generally, the process model with the highest fitness value can cover more process instances. For example, in Case G6, the process model derived from the population of 500 generations has the good fitness value of 0.9118. It represented 133 process instances. But, Case G1 only represented 46 processes, because it comes from the population of 50 generations and has a lower fitness value of 0.8355 that Case G6. The process model with higher fittest value generally represents more process instances. However, the relationship between the amount of generations and the coverage value does not fully follow this pattern. It means that the coverage value does not necessarily increase with an increasing number of generations.

The second group of tests was designed to examine the impact of population size on process mining. The results illustrate that both average fitness and best fitness increased as population size increased (Table 7.2). However, Case P1 was an exception. Compared with Case P2 and P3, P1 had a higher average fitness. Although the highest process coverage was held by the process model with the highest fitness, the relationships among the size of population, the value of average fitness, and the coverage of process instances were not strong enough. For example, the process model in Case P2 was derived from a population with a lower average fitness and lower fittest, but it represented more process instances than other process model except the one in Case P6. The model is an abstraction of a dynamic system, and the purpose of process miming is to discover a process model that can represent the behavior of a system well. Therefore, the best process model is expected to parse as many instances as possible process in the event log.

The third group of tests was designed to examine the impact of crossover rate on process mining. The results illustrate that the relationships among the average fitness of population, the best fitness of individuals, the coverage of process instances, and the rate of crossover were very complex (Table 7.2). When the rate of crossover increased from 0.19 to 0.59, both the best fitness and process instances coverage increased. When crossover rate increased from 0.59 to 0.79, both the best fitness and process instances coverage decreased. The average fitness generally increased as the crossover rate increased, but there was an exception (*e.g.*, when crossover rate equaled 0.79).

The fourth group of tests was designed to examine the impact of mutation rate on process mining. The results show that the relationships among the average fitness of a population, the best fitness of individuals, the coverage of process instances, and the rate of mutation were very complex also (Table 7.2). When the mutation rate increased from 0.005 to 0.05, the average fitness, the best fitness, and the process instances coverage increased except Case M1. Although Case M1 had the highest average fitness (0.9642) and the highest process instances coverage (221), the population was not introduced any new individuals through the operation of mutation, because with the mutation of 0.005, only a half individual can be introduced into the population with 100 individuals. M1 could be a typical example of local optimization. In general, the mutation rate increased from 0.05 to 0.30, the average fitness, the best fitness, and the process instances coverage decreased.

The fifth group of tests was designed to examine the impact of elitism rate on process mining. The results suggest that the relationships among the average fitness of population, the best fitness of individuals, the coverage of process instances, and the elitism rate were not strong enough (Table 7.2). The average fitness and best fitness generally increased when the rate of elitism increased except when elitism rate was 0.02. The process instance coverage increased when the elitism rate increased from 0.01 to 0.04, and it decreased when the elitism rate continued to increase from 0.04 to 0.40.

7.2.2. Urban Growth Process

Based on the tests above, the optimized GA parameters configuration was developed as follows: the number of generations 500, population size 300, crossover rate 0.59, mutation rate 0.05, elitism rate 0.04, and pruning rate 0.02. The last group tests in this experiment were about LCLU change processes. Urban growth was treated as an example. The study dealt with a set of 6 time series data with 7 LCLU features. Therefore, theoretically, a total of 7^6 (or 117649) possible process models can be defined. Those process models can be categorized into 7^2 (or 49) groups based on the start states and end states of LCLU features. For each group, there were 7^4 (2401) possible process models. The results of the sixth group tests were given as follows.

Figure 7.5a shows the URBAN to URBAN process models during 1990-2002. Obviously, URBAN => URBAN => URBAN => URBAN => URBAN => URBAN should be the major process of URBAN to URBAN process models group. Urban areas that follow this process model did not change. Due to the complexity of urban feature and noise data, the pruning rate was set at 10%. Several unlikely processes were deleted, such as, URBAN => FARM => FARM => FARM => URBAN => URBAN, and URBAN => WETLAND => WETLAND => FARM => URBAN => URBAN. Those process models were unlikely in the real world. A total of 411 process instances were taken and used in the event log file, and 325 were represented by this process model. It has been mentioned that 2401 possible process models can be defined in this group. Therefore, 79.08% process instances were represented by 0.0416% model.

Figure 7.5b shows the FARM to URBAN process models during 1990-2002. After performing 5% pruning , only three process models remained in this group. They are FARM => GRASS => GRASS => URBAN => URBAN => URBAN, FARM => FARM => FARM =>

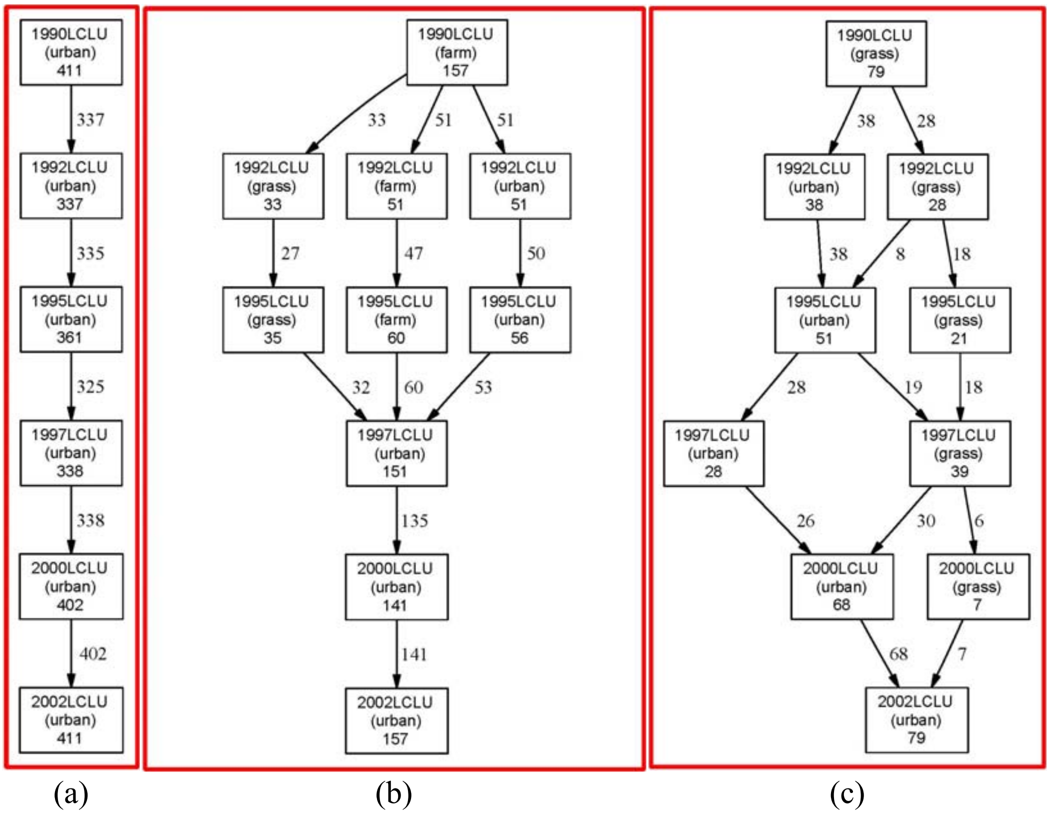
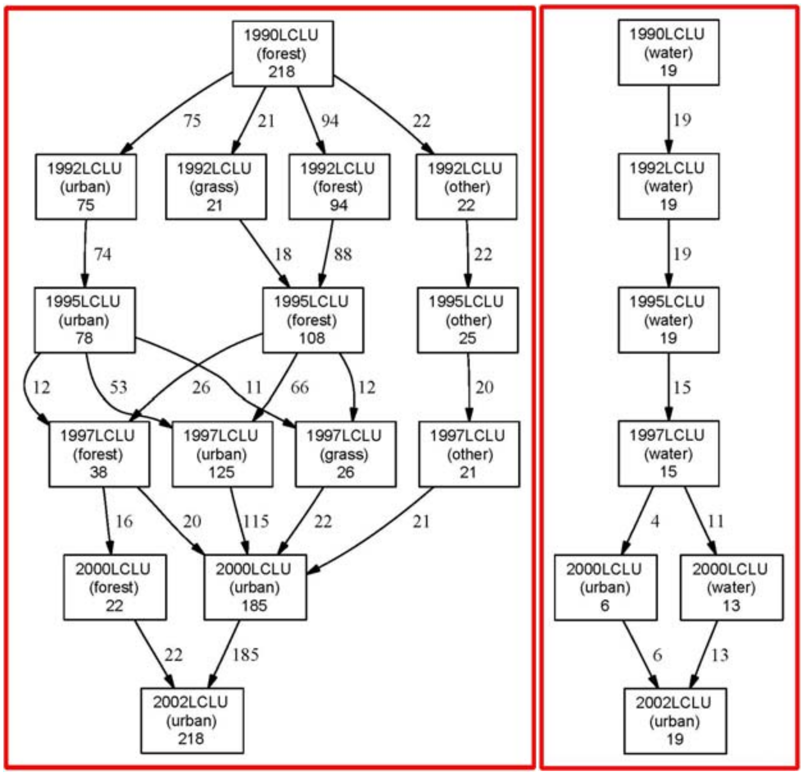
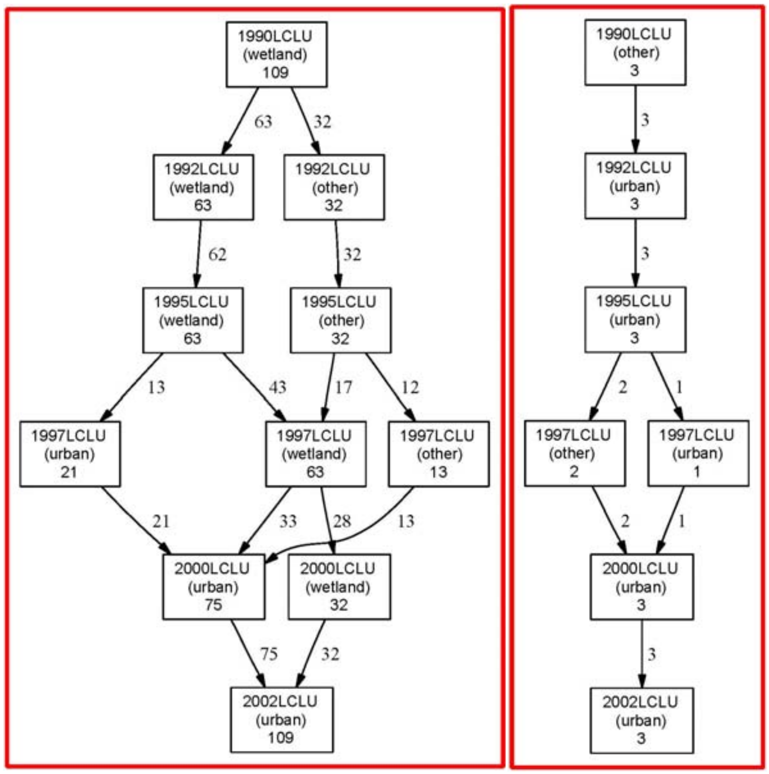


Figure 7.5. Process models representing land-cover and land-use change (from (a) URBAN to URBAN, (b) FARM to URBAN, and (c) GRASS to URBAN)



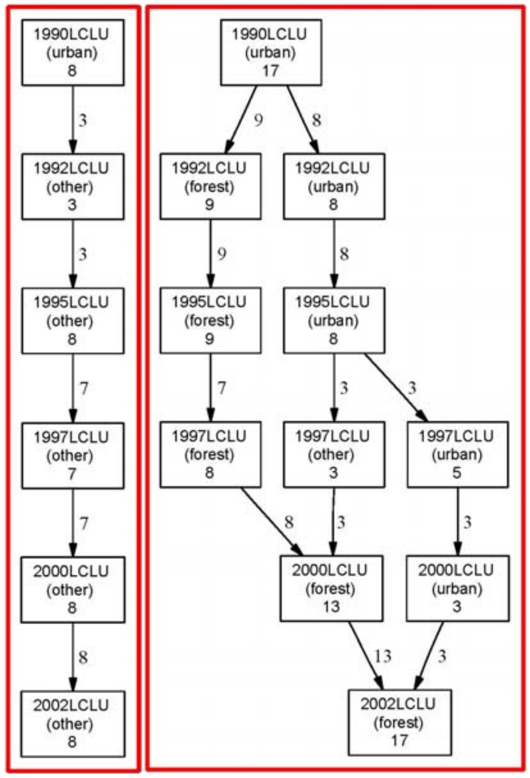
(d) (e)

Figure 7.5. (Continued) (from (d) FOREST to URBAN and (e) WATER to URBAN)



(f) (g)

Figure 7.5. (Continued) (from (f) WETLAND to URBAN and (g) OTHER to URBAN)



(h) (i)

Figure 7.5. (Continued) (from (h) URBAN to GRASS and (i) URBAN to FOREST)

URBAN => URBAN => URBAN, and FARM => URBAN => URBAN => URBAN => URBAN => URBAN. These three are only 0.1249% of total process models, but they can parse 78.98% process instances.

Figure 7.5c shows the GRASS to URBAN process models during 1990-2002. After performing 6% pruning, only seven process models remained in this group (Table 7.4). Again, theoretically, there are 2401 possible process models in this group. The remaining process models are only 0.3332% of them, but these process models can parse 78.48% process instances.

One of the most important process groups is FOREST to URBAN (Figure 7.5d). Obviously, according to the results, the amount of process models was not as many as the theoretical amount of 2401. The unlikely process models, such as, FOREST => URBAN => URBAN => GRASS => FARM => URBAN, and FOREST => FARM => BARE => WATER => FARM => URBAN, should be eliminated. After performing 4% pruning, the most unlikely process models had been eliminated, but the process models with low possibility, such as, FOREST => URBAN => URBAN => FOREST => FOREST => URBAN, and FOREST => URBAN => URBAN => GRASS => URBAN => URBAN, were still remained (Figure 7.4). All remained process models were interpreted and illustrated as in Table 7.3. To discover the process models of the LCLU change from FOREST to URBAN during 1990 to 2002, 218 process instances were sampled and listed in the event log file. After pruning, the final process models on the diagram were interpreted into 13 individual process models on the table. Although those 13 process models were only 0.0054% of 2401 theoretical possible process models, they can parse 194 process instances, which is 88.99% of total 218 process instances.

Figure 7.5e shows the WATER to URBAN process models during 1990-2002. This group of change processes occur when man-made structures were located near or over water bodies, such as, bridges and docks. After pruning 10%, only two final process models remained:

WATER => WATER => WATER => WATER => WATER => URBAN, and WATER => WATER => WATER => WATER => URBAN => URBAN. They represented 78.95% of WATER to URBAN process instances.

Figure 7.5f shows the WETLAND to URBAN process models during 1990-2002. After performing 7% pruning, only six process models remained in this group (Table 7.4). They are only 0.2499% of theoretical possible process models. They represented 85 process instances that is 77.98% of total sampling process instances.

Figure 7.5g shows the OTHER to URBAN process models during 1990-2002. This group of change processes occurs when OTHER features were used for urban development, such as factory or any construction. After pruning 2%, there were only two final process models remained: OTHER => URBAN => URBAN => URBAN => URBAN => URBAN, and OTHER => URBAN => URBAN => GRASS => URBAN => URBAN. Because it was a small feature, two process models represented 100% of OTHER to URBAN process instances.

Figure 7.5h shows the URBAN to GRASS and Figure 7.5i shows URBAN to FOREST process models during 1990-2002. Both groups of change processes likely occur in the residential area and require a long time. For example, after developing a residential area, grass, shrub, and trees are usually planted soon. Though trees require a long time to grow, grass and shrub grow fast. After pruning respectively, the final process model represented 37.50 % of URBAN to GRASS process instances and 82.35% of URBAN to FOREST process instances, respectively. In this study, the areas following the process models of URBAN to GRASS and URBAN to FOREST will be recognized as the planted URBAN areas.

7.2.3. Process Models Accuracy Assessment

Process mining is a new technique. Although it has been used in business and manufacturing management, this dissertation research is the first attempt in GIS and remote sensing area. There is no standard method to examine the accuracy of the discovered LCLU change processes. Here, the accuracies of LCLU change processes were assessed by using average fitness, best fitness, and process instances coverage first (Table 7.3). After evaluating the discovered process, models used the average fitness, the best fitness, and the coverage of process instances. A total of 38 process models for urban growth were selected. These process models were translated from Petri net (PN) into individual process models (Table 7.4). These process models can be used as cellular automata (CA) transition rules in spatio-temporal modeling in the next chapter.

The second part of accuracy assessment was done by evaluating the agreement between the end year LCLU and the re-projected end year LCLU. The average fitness, the best fitness, and the coverage of process instances were used to assess the accuracy of the process model directly. The agreement between the end year LCLU and the re-projected end year LCLU can be used as an indirect accuracy assessment for the process models. The assumptions were: 1) if a process model was developed from a population with high average fitness, high fittest, and high instances coverage, then this process model could be regarded as a potentially good one; 2) the process models with good accuracy can be used to create a re-projected LCLU image with a good accuracy.

Due to the complexity of the study, only the urban area was re-projected by coding those process models into the expert system in ERDAS Imagine 9.0. The resultant image was used to compare with the LCLU in 2002 (Figure 7.6). The study area was covered by 4,257,000 pixels, and among them, 134,114 pixels were classified as urban in 2002. The re-projected urban area for the same year had 124,726 pixels. The agreement between the two images was 90.01%. There is still some room to improve. The process models with high accuracy level were selected for cellular automata based study in the next chapter.

Table 7.3. Urban growth process mining*

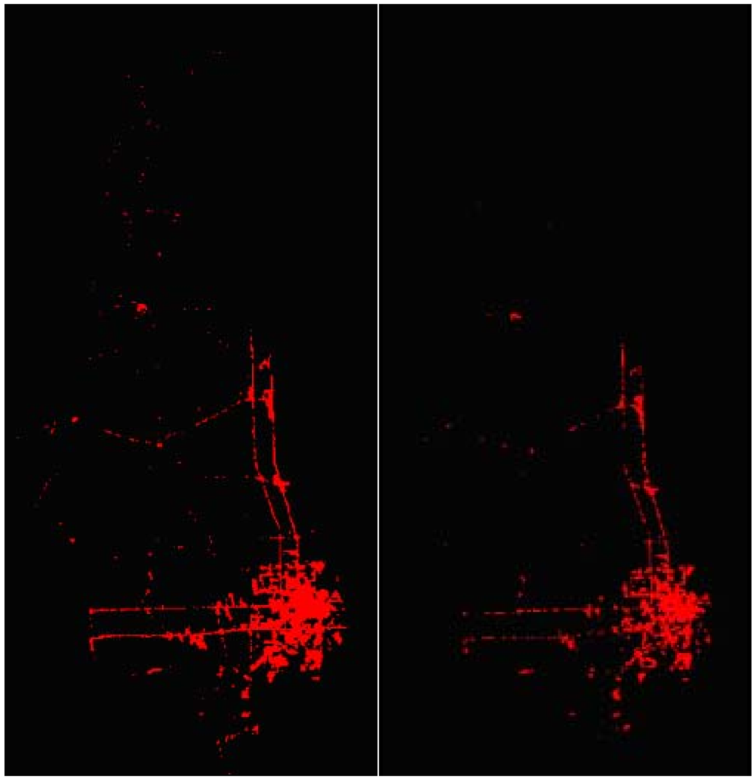
Start States	End States	Average Fitness	Best Fitness	Sampling Instances	Covering Instances	Coverage Percentage	Number of Models	Percentage of Models
Urban	Urban	0.9539	0.9848	411	325	79.08%	1	0.0416%
Farm	Urban	0.9551	0.9788	157	124	78.98%	3	0.1249%
Grass	Urban	0.9535	0.9842	79	62	78.48%	7	0.3332%
Forest	Urban	0.9511	0.9810	218	194	88.99%	13	0.0054%
Water	Urban	0.9584	0.9822	19	15	78.95%	2	0.0833%
Wetland	Urban	0.9657	0.9854	109	85	77.98%	6	0.2499%
Other	Urban	0.9518	0.9846	3	3	100.00%	2	0.8330%
Urban	Grass	0.9520	0.9834	8	3	37.50%	1	0.0416%
Urban	Forest	0.9527	0.9841	17	14	82.35%	3	0.1249%
Total		--	--	1021	825	80.80%	38	0.1759%

* The average fitness and the best fitness were based on the un-pruned Petri nets. The instances coverage, coverage percentage, and percentage of models were based on the pruned Petri nets.

Table 7.4. Top 38 models of urban growth processes after pruning*

Processes	1990	1992	1995	1997	2000	2002
1	Urban	Urban	Urban	Urban	Urban	Urban
2	Farm	Grass	Grass	Urban	Urban	Urban
3	Farm	Urban	Urban	Urban	Urban	Urban
4	Farm	Farm	Farm	Urban	Urban	Urban
5	Grass	Urban	Urban	Urban	Urban	Urban
6	Grass	Urban	Urban	Grass	Urban	Urban
7	Grass	Urban	Urban	Grass	Grass	Urban
8	Grass	Grass	Grass	Grass	Grass	Urban
9	Grass	Grass	Urban	Urban	Urban	Urban
10	Grass	Grass	Urban	Grass	Grass	Urban
11	Grass	Grass	Urban	Grass	Urban	Urban
12	Forest	Urban	Urban	Forest	Forest	Urban
13	Forest	Urban	Urban	Forest	Urban	Urban
14	Forest	Urban	Urban	Urban	Urban	Urban
15	Forest	Urban	Urban	Grass	Urban	Urban
16	Forest	Grass	Forest	Forest	Forest	Urban
17	Forest	Grass	Forest	Forest	Urban	Urban
18	Forest	Grass	Forest	Urban	Urban	Urban
19	Forest	Grass	Forest	Grass	Urban	Urban
20	Forest	Forest	Forest	Forest	Forest	Urban
21	Forest	Forest	Forest	Forest	Urban	Urban
22	Forest	Forest	Forest	Urban	Urban	Urban
23	Forest	Forest	Forest	Grass	Urban	Urban
24	Forest	Other	Other	Other	Urban	Urban
25	Water	Water	Water	Water	Water	Urban
26	Water	Water	Water	Water	Urban	Urban
27	Wetland	Wetland	Wetland	Urban	Urban	Urban
28	Wetland	Wetland	Wetland	Wetland	Urban	Urban
29	Wetland	Wetland	Wetland	Wetland	Wetland	Urban
30	Wetland	Other	Other	Wetland	Urban	Urban
31	Wetland	Other	Other	Wetland	Wetland	Urban
32	Wetland	Other	Other	Other	Urban	Urban
33	Other	Urban	Urban	Urban	Urban	Urban
34	Other	Urban	Urban	Grass	Urban	Urban
35	Urban	Grass	Grass	Grass	Grass	Grass
36	Urban	Forest	Forest	Forest	Forest	Forest
37	Urban	Urban	Urban	Urban	Urban	Forest
38	Urban	Urban	Urban	Grass	Forest	Forest

*Areas following URBAN to GRASS or to FOREST will be still regarded as the planted URBAN areas.



(a) (b)

Figure 7.6. Urban area in 2002: (a) classified image and (b) modeled image

7.3. Discussion

7.3.1. Impact of Genetic Parameters on Land-Cover and Land-Use Change Process Mining

The impacts of genetic parameters on the performance of GA-based Petri net includes at least four aspects: 1) premature convergences, 2) local optimization, 3) not enough evolution, and 4) unstable population. The improper use of GAs may result in low-quality process models. It is very important to examine the relationships among genetic parameters, the performance of GAs, and the quality of the resultant process models.

Convergence can be regarded as the state when most of the population is identical or diversity is minimal (Louis and Rawline, 1993). Premature convergence is a phenomenon in which population cannot converge to a better level of fitness because the better individuals are eliminated (Cheng *et al*., 1996). One of the cases shown in Table 7.2 is premature convergence. Case C1 had the lowest crossover rate at 0.09. Crossover occurs between different individuals. When the crossover rate was set at a low level, the diversity of population will be limited. The crossover between similar types of individuals will generate offsprings that have similar fitness and capability, so that the evolution of a population will converge to a low level, and the best individual's capability will be limited. In Case C1, the average fitness and the best fitness of the population were 0.3920 and 0.8768, respectively. The best process models group can only represent 57 process instances, which is only 26.15% of total samples.

Optimization is the process to find the best solution. It usually has two levels: local and global. The local optimization solution can be good only at a particular space or time. The global optimization solution should be good anywhere and anytime. Local optimization in GAs is usually achieved by improper configuration of genetic parameters (*e.g.*, high elitism rate 0.40). Case E6 in this experiment is an example of local optimization. The high elitism rate swaps many high fitness individuals into the population. After a number of generations, the population was dominated by elites. In this case, both average fitness and best fitness were usually high (*e.g.*, 0.6910 and 0.9267, respectively), but the performance of process mining was very poor in terms of instance coverage. The problem is that the diversity of the population is low, and the process models developed in this situation cannot parse most process instances. In Case E6, the process instances coverage was 78, which is only 35.78% of the total samples.

Unstable population is a phenomenon in which a population cannot hold a structure for Long-term evolution. The high fitness genome cannot be kept and transferred to the next generation because they were destroyed. One of the cases shown in Table 7.2 has an unstable population problem. Case M6 had the highest mutation rate of 0.30. When the mutation rate is high, the random search will dominate the behavior of GA, and the optimization capability will be lost. Hence, the average fitness and best fitness values were lower than the rest of the cases (Table 7.2). When the performance of GAs is poor, the quality of the process model will be poor also.

7.3.2. Complexity of Process Models

Spatio-temporal process mining is a very complex task that is impacted by many factors. LCLU classification is the first factor that may impact process mining significantly. If the quality of LCLU classification is not good enough, the quality of the event log file will be affected. Obviously, the quality of resultant process models will be a problem also. Therefore, both LCLU classification and change detection play a very important role in developing the process models.

Even though no major problem appears during the phase of LCLU classification, event log file editing, and process models generation, the resultant processes were complex. In this research, we dealt with 7 LCLU features for 6 time periods. Theoretically, a total of 7^6 (or 117649) possible process models can be defined. Those process models can be categorized into 7^2 (or 49) groups based on the start states and end states of LCLU features. Each group, theoretically, had 7^4 (or 2401) possible process models. After pruning, there were still many process models remained in the system. For example, the group of FOREST to URBAN process had 13 process models, and the group of GRASS to URBAN process had 7 process models (Table 7.4).

Each process model depicts changes both spatially and temporally. Although different pixels may share same process model, many different pixels have different process models. Based on the observation of the process models, such as, Figure 7.4b, the process of FARM to FARM was the same during 1990-1992 and 1992-1995, but the transition rate was different. Process models provide the insight into the LCLU change processes. They also reveal the complexity of spatio-temporal process mining.

7.3.3. Accuracy Assessment of Process Models

Because there are no standard techniques for process model accuracy assessment, two approaches were used to assess the accuracy of process models. The first approach was based on the average fitness, the best fitness, and the coverage of process instances. The second approach was based on the agreement between the end year LCLU and re-projected end year LCLU. In particular, only the urban feature was re-projected and assessed in this chapter. These two approaches are complementary to each. When the process models have high value of average fitness of population and best fitness of individual, their coverage of process instances will likely be high also. The process models will likely perform well in terms of representing the processes in the real data, and the agreement between the end year LCLU and the re-projected year LCLU will be high.

7.4. Hypothesis #4 Review

The fourth research question as presented in Chapter 1 was how do genetic parameters (such as number of generations, population size, crossover rate, mutation rate, and generation gap) impact the accuracy of process-oriented change detection. The fourth research hypothesis presented in Chapter 1 was stated as follows:

Ha: Different number of generations, population size, crossover rate, mutation rate, or generations gap can increase / decrease the accuracy of genetic PN-based LCLU change process detection, respectively.

Based on the experiment designed and performed in this chapter, genetic parameters can affect the performance of genetic PN-based LCLU change detection significantly. In the case of the number of generations, when the number of generations is small (*e.g.*, less than 50 generations), the fitness of genetic PN-based LCLU change process models was low. The fitness significantly increased as the number of generations increased from 50 to 500.

In the case of population size, although a larger population theoretically leads to good fitness, the relationships among the size of population, the value of average fitness, and the coverage of process instances were not strong enough. For example, the process model in Case P2 was derived from a population with a lower average fitness and lower best fitness, but it represented more process instances than many other process models.

In the case of crossover rate, the results illustrate that the relationships among the average fitness of population, the best fitness of individual, the coverage of process instances, and the rate of crossover were very complex. When the rate of crossover was small (*e.g.*, less than 59%), the accuracy of GA-based LCLU classification increased. Whereas it was larger than 0.59, the best fitness and process instances coverage decreased.

In the case of mutation rate, the results show that the relationships among the average fitness of population, the best fitness of individual, the coverage of process instances, and the rate of mutation were very complex also. When the mutation rate increased from 0.005 to 0.05, both the best fitness and the process instances coverage increased except Case M1. When the mutation rate increased from 0.05 to 0.30, both the best fitness and the process instances coverage decreased.

Finally, in the case of elitism rate, the results suggest that the relationships among the average fitness of population, the best fitness of individual, the coverage of process instances, and the elitism rate were not strong. The average fitness and best fitness generally increased when the rate of elitism increased except when the elitism rate was 0.02. The process instance coverage increased when the elitism rate increased from 0.01 to 0.04, and it decreased when the elitism rate continued to increase from 0.04 to 0.40.

The impacts of various genetic parameters on the performance of GA-Petri net-based LCLU change detection, as described above, are various. In general, the performance can be improved by increasing the number of generations and population size. In order to improve the performance, crossover rate, mutation rate, and elitism rate should be kept at middle level.

7.5. Summary

Although the primary goal of this study is to investigate the relationship between genetic parameters and genetic PN-based LCLU change detection, the LCLU change process mining and accuracy was also examined. The relationships among the factors affecting the performance of genetic PN-based LCLU change detection are very complex, and the accuracy assessment of the LCLU process models is still a great challenge. The major conclusions can be made as follows:

- GA parameters can significantly impact the performance of genetic PN-based LCLU change detection, especially the LCLU change process mining. When genetic parameters are set improperly, the average fitness of population, the best fitness of individual, and the coverage of process instances will be low.
- In order to improve the performance of genetic PN-based LCLU change detection, the number of generations and population size should be increased, and the crossover

rate, mutation rate, and elitism rate should be kept at a middle level. A recommended configuration includes: generation 500, population 300, crossover 59%, mutation 5%, and elitism rate 4%.

- The study has demonstrated the usefulness of genetic PN-based LCLU change detection. It can provide information about what, when, why, how, and how much LCLU change. The process models can be used as transition rules in CA-based spatio-temporal modeling.

CHAPTER 8 IMPACT OF CELLULAR AUTOMATA COMPONENTS ON LAND-COVER AND LAND-USE PREDICTIVE MODELING

8.1. Experiment #5 Design

If process-oriented LCLU change detection is regarded as the procedure of turning patterns into processes, predictive modeling is regarded as a procedure of turning the current processes into future patterns. In Chapter 7, the major issue examined was about genetic PN-based LCLU change processes detection, and the goal was to find the LCLU change processes. The major issue in this chapter is about cellular automata (CA)-based LCLU predictive modeling, and the goal was to predict future LCLU patterns and reconstruct past LCLU patterns. The importance of LCLU modeling is to predict future LCLU and reconstruct past LCLU. The future LCLU information can be used for regional development planning, and the past LCLU information can be used for historical LCLU research. Both of them can be used to improve the understanding of human-nature relationship.

This experiment was designed to examine the relationships between CA components and the performance of predictive modeling, especially to find out better CA-component configurations for predicting the future LCLU and reconstructing the past LCLU. Six images were acquired during 1990-2002 (Figure 3.1). The AOIs of these images were classified (Table 7.1 and Figure 7.1), and the change process models were discovered based on the classified images. The experiments done in the previous chapters have become the foundation of the experiment in this chapter. Experiment #5 had six groups of tests. Among them, five groups of tests were used to find the optimized CA components, and one group of tests was used to predict the future LCLU or reconstruct the past LCLU (Figure 8.1). The model calibration was performed by: 1) evaluating the agreement between predicted 2005 LCLU and classified 2005 LCLU, and 2) adjusting the configuration of CA components.

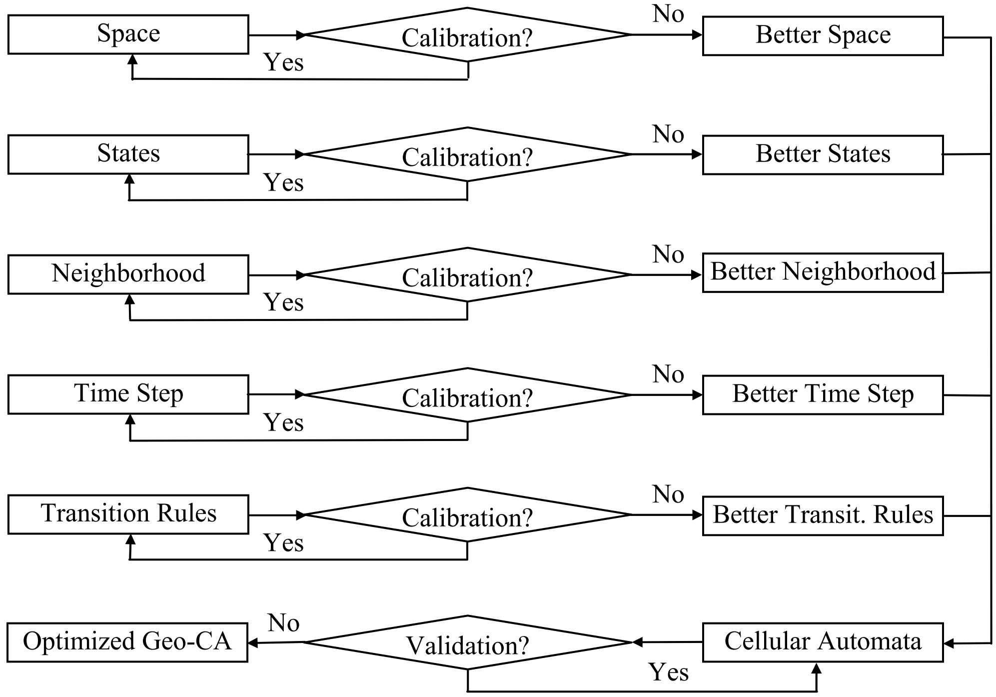


Figure 8.1. The workflow of CA-based land-cover and land-use predictive modeling

The first group of tests was designed to examine the impact of space size on the performance of CA-based modeling. Three space sizes tested were 6025 x 12993, 1507 x 3249, and 378 x 813 pixels. These space sizes correspond to spatial resolutions of 7.5m x 7.5m, 30m x 30m, and 120m x 120m, respectively.

The second group of tests was designed to examine the impact of number of states on the performance of CA-based modeling. Only two sets of states were tested: 1) urban and non-urban, and 2) urban, farm, grass, forest, water, wetland, and other.

The third group of tests was designed to examine the impact of neighborhood type and size on the performance of CA-based modeling. Two types and two sizes of neighborhood were tested: 1) Moore neighborhood 3 x 3, 2) extended Moore neighborhood 7 x 7, 3) von Neumann neighborhood 3 x 3, and 4) extended von Neumann neighborhood 7 x 7.

The fourth group of tests was designed to examine the impact of time steps on the performance of CA-based modeling. Based on the availability of data sources, only two time steps were examined: 2 - 3 years, and 4 - 5 years.

The fifth group of tests was designed to examine the impact of transition rules on the performance of CA-based modeling. The transition rules of CA are the most important components. They control the status of each cell. In this experiment, the transition rules were developed based on the process models. According to the experiment in Chapter 7, the process models discovered under different genetic configurations have different qualities. In this group of tests, only two sets of transition rules were tested. They were developed by different genetic configurations. The first genetic configuration was: generation 500, population 300, crossover 0.59, mutation 0.05, and elitism 0.04. The second genetic configuration is generation 50, population 50, crossover 0.09, mutation 0.01, and elitism 0.01. The pruning rate was 1% - 10% for both of them.

The last group of tests was designed to perform future LCLU prediction. The configuration of CA components was set up based on previous five groups of tests. In fact, the goal of those tests is to calibrate the CA models. The modeling-based 2002 LCLU maps were compared with the classification-based 2002 LCLU map. Based on the agreement between them, the configuration of CA components was adjusted. The procedure was repeated until good agreement was achieved. After calibration, the CA model was validated by comparing prediction-based 2005 LCLU and classification-based 2005 LCLU, and the agreement between them was evaluated.

8.2. Results

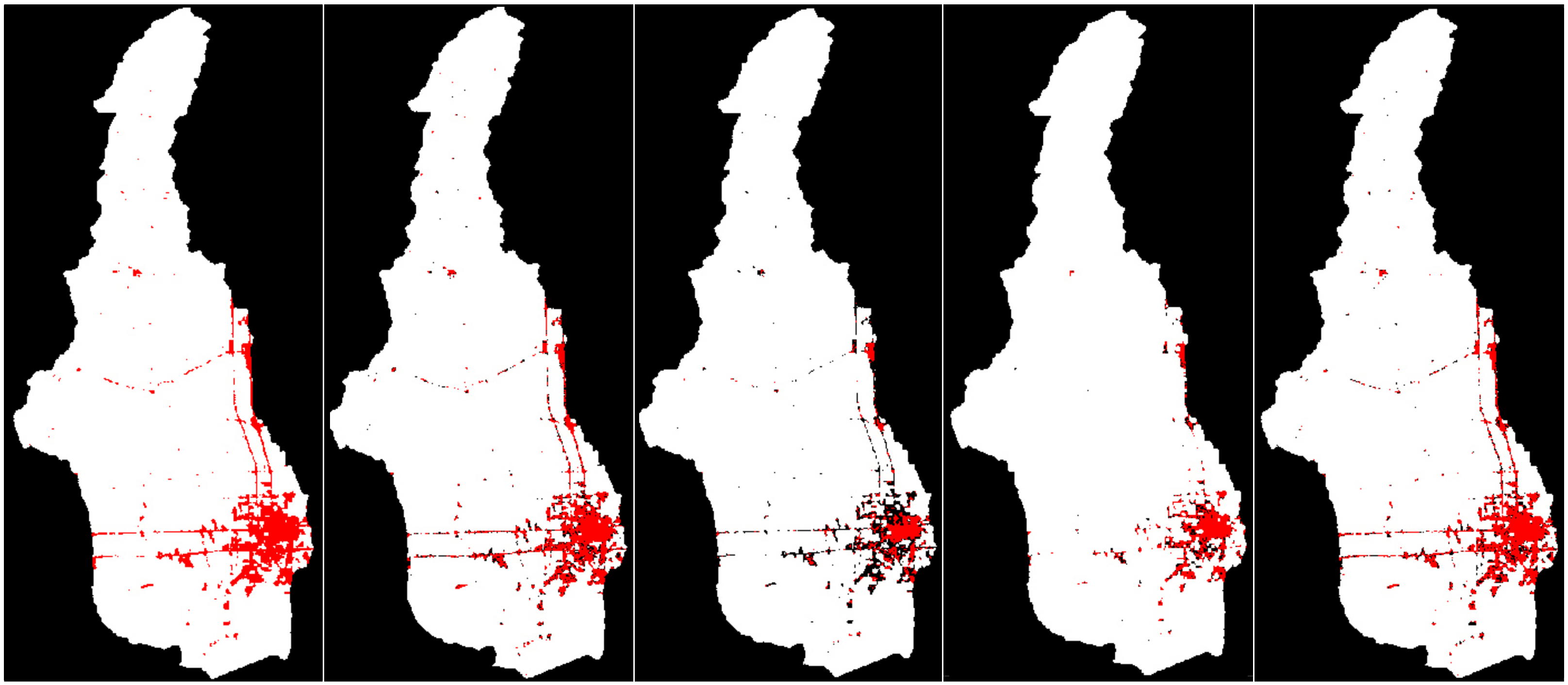
This research focuses only on the major issues in the CA-based LCLU predictive modeling. These issues include the configuration of CA components, the calibration of CA-based LCLU models, and future LCLU prediction. The default settings of CA components were: 1) space – 1507 x 3249 pixels, 2) states – 7 classes, 3) time step – 2.5 years, 4) neighborhood – 3 x 3 Moore neighborhood, 5) transition rules – developed under the condition of optimized GA configuration. When a component was changed and tested, other components remained at their default values. Theoretically, all LCLU features should be involved in the procedure of calibration. Because calibration is a time-consuming procedure and URBAN is the most complex LCLU feature, only URBAN was used for the calibration purpose in this dissertation research. After identifying the optimized configuration of CA components, future LCLU prediction was performed. Currently, there is no standard method for calibrating CA-based spatio-temporal models, the traditional LCLU classification accuracy assessment was used for calibrating CA-based modeling. The idea is to compare the modeling resultant images derived from different CA configurations with the classified image. The classified image was used as a reference. The best model should have the highest agreement between the classified image and the modeled image.

Table 8.1. Calibration of CA-based land-cover and land-use predictive modeling*

Components of Cellular Automata		Case ID**	Setting	Tested Total	Number Correct	Producer Accuracy	User Accuracy	Overall Accuracy	Kappa Value
Space (cells)		B	378 x 813	500	155 298	77.50% 99.33%	98.73% 92.26%	90.60%	0.8069
		F	1507 x 3249	500	133 300	66.50% 100.00%	100.00% 90.91%	86.60%	0.7307
		E	6025 x 12993	500	165 300	82.50% 100.00%	100.00% 100.00%	93.00%	0.8622
States (classes)		G	2	500	179 300	85.00% 100.00%	100.00% 91.19%	94.00%	0.8721
		F	7	500	133 300	66.50% 100.00%	100.00% 90.91%	86.60%	0.7307
Time Steps (years)		F	2 - 3	500	133 300	66.50% 100.00%	100.00% 90.91%	86.60%	0.7307
		H	4 - 5	500	116 300	58.00% 100.00%	100.00% 90.91%	83.20%	0.6714
Neighborhood	Moore	F	3 x 3	500	133 300	66.50% 100.00%	100.00% 90.91%	86.60%	0.7307
		D	7 x 7	500	98 300	49.00% 100.00%	100.00% 77.12%	79.60%	0.5515
	von Neumann	I	3 x 3	500	144 300	72.00% 100.00%	100.00% 92.88%	88.80%	0.7747
		J	7 x 7	500	116 300	58.00% 100.00%	100.00% 82.64%	83.20%	0.6438
Transition Rules		C	Un-optimized	500	67 300	33.50% 100.00%	100.00% 90.91%	73.40%	0.5167
		F	Optimized	500	133 300	66.50% 100.00%	100.00% 90.91%	86.60%	0.7307

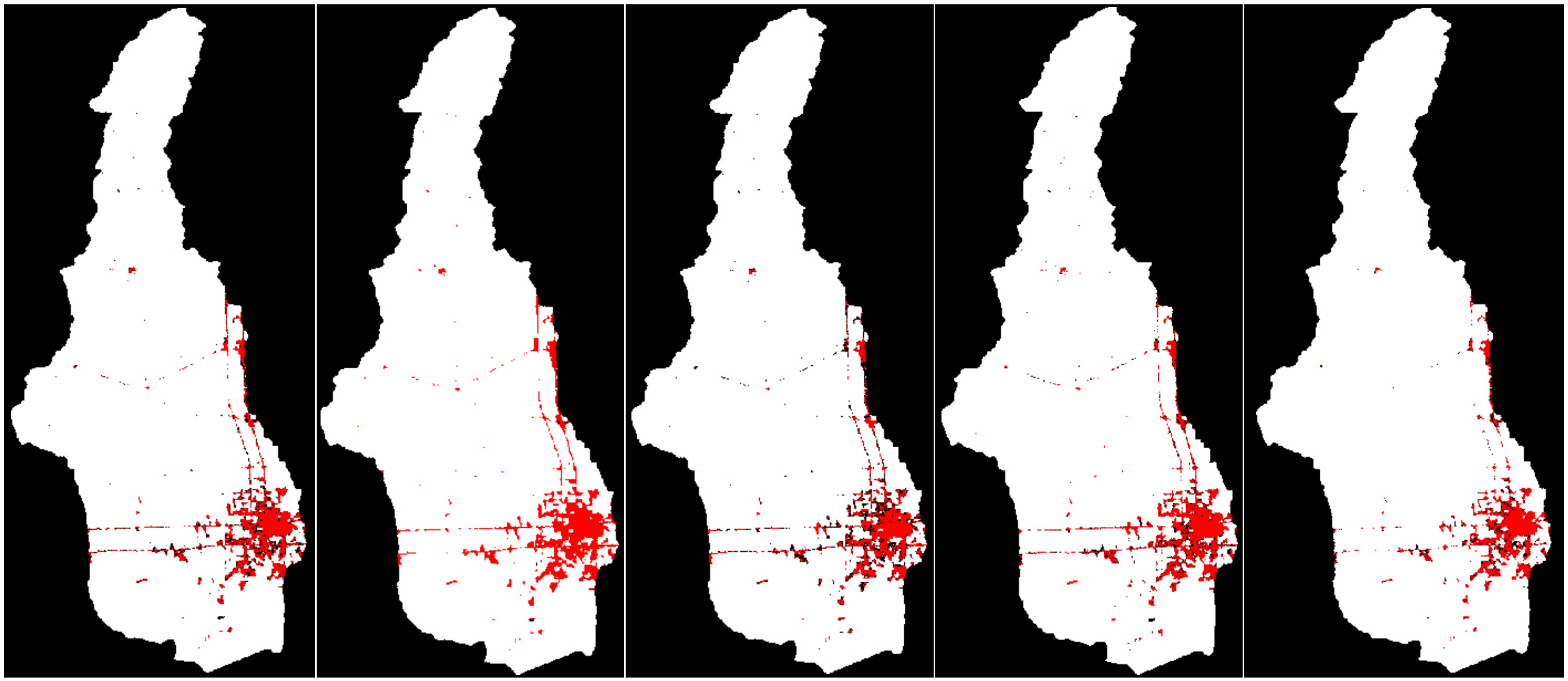
* For Number Correct, Producer Accuracy, and User Accuracy, the first number is about URBAN, and the second number is NON-URBAN. The calibration was done by comparing classified result and modeling result, and it was based on URBAN and NON-URBAN.

** F is the case with default CA configuration: space - 1507x3249, state - 7, time step – 2 to 3 years, neighborhood - Moore (3x3), and transition rule - optimized.



A. 2002 Classified urban B. Space 378 x 813 C. Un-optim. transition D. Moore 7x7 E. Space 6025 x 12993

Figure 8.2. Calibration of CA-based land-cover and land-use modeling – taking URBAN as an example



F. Default setting | G. States 2 classes | H. Time step 5 years | I. von Neumann 3 x 3 | J. von Neumann 7 x 7

Figure 8.2. (Continued)

8.2.1. Space

The space of CA is composed of individual cells. The cell is the basic spatial unit, and it has two characteristics: shape and size. The square shape of cell is similar to the shape of satellite image pixel. In order to predict URBAN growth in the study area, the square shape of cell was adopted. Three space sizes were used in the tests: 378 x 813, 1507 x 3249, and 6025 x 12993. The results indicate that the overall accuracy changed from 90.60% to 86.60% and then to 93.00%, and the overall kappa value changed from 0.8069 to 0.7307 and then to 0.8622 when the space size increased from 378 x 813 to 1507 x 3249, and then to 6025 x 12993 (Table 8.1). Figure 8.2 demonstrates the comparison among 2002 classified image and modeling images derived from various CA space configurations. Images B, D, and E represent the results corresponding to space sizes 378 x 813, 1507 x 3249, and 6025 x 12993 respectively.

8.2.2. States

The state of cell in a cellular automaton may represent any spatial variable, for example, the various types of LCLU. The LCLU in the study area was classified into seven classes, and each of them represented a cell state. The state of a cell may be dynamic or static. For example, the state of cell changes from grass land to farm land, or remains water. However, even for the water body, the change can still happen. When building a bridge over a water body, the state of cell changes from water to urban. In this dissertation study, the state of all cells was assumed as dynamic, and the impact of different number of state on the performance of CA-based LCLU modeling was tested. In order to examine the impact of number of state on the performance of CA-based LCLU modeling, two sets of data were used. The first set of data has two states: urban and non-urban. The second set of data has seven states: urban, farm, grass, forest, water, wetland, and other. The results indicate that the overall accuracy decreases from 94.00% to 86.60% and the overall kappa value changes from 0.8721 to 0.7307 when the number of state changes from 2 to 7 (Table 8.1). Figure 8.2 compares the 2002 classified image and modeled LCLUs derived from various CA state configurations. Images E and F represent the results corresponding to states number 7 and 2 respectively.

8.2.3. Neighborhood

The neighborhood of a cell is defined as a set of neighboring cells surrounding the interested cell. The classical CA has two types of neighborhood: von Neumann and Moore. Usually, four neighboring cells will be considered as neighbor in von Neumann neighborhood, and eight neighboring cells will be considered as neighbor in Moore neighborhood. In addition to these two types of neighborhood, four more neighborhoods were also tested. They were created by extending the radius of von Neumann or Moore from 3 x 3 to 7 x 7. The results indicate that the overall accuracy changed from 86.60% to 79.60% and overall kappa value decreased from 0.7307 to 0.5515 when the size of Moore neighborhood increased from 3 x 3 to 7 x 7 (Table 8.1). The results also indicate that the overall accuracy changed from 88.80% to 83.20% and overall kappa value decreased from 0.7747 to 0.6438 when the size of the von Neumann neighborhood increased from 3 x 3 to 7 x 7 (Table 8.1). Overall, the von Neumann neighborhood yielded more accurate results than the Moore neighborhood. Figure 8.2 compares the 2002 classified image and modeled LCLUs derived from various CA neighborhood configurations. Images F, D, I, and J represent the results corresponding to neighborhoods Moore 3 x 3, Moore 7 x 7, von Neumann 3 x 3, and von Neumann 7 x 7, respectively.

8.2.4. Time Steps

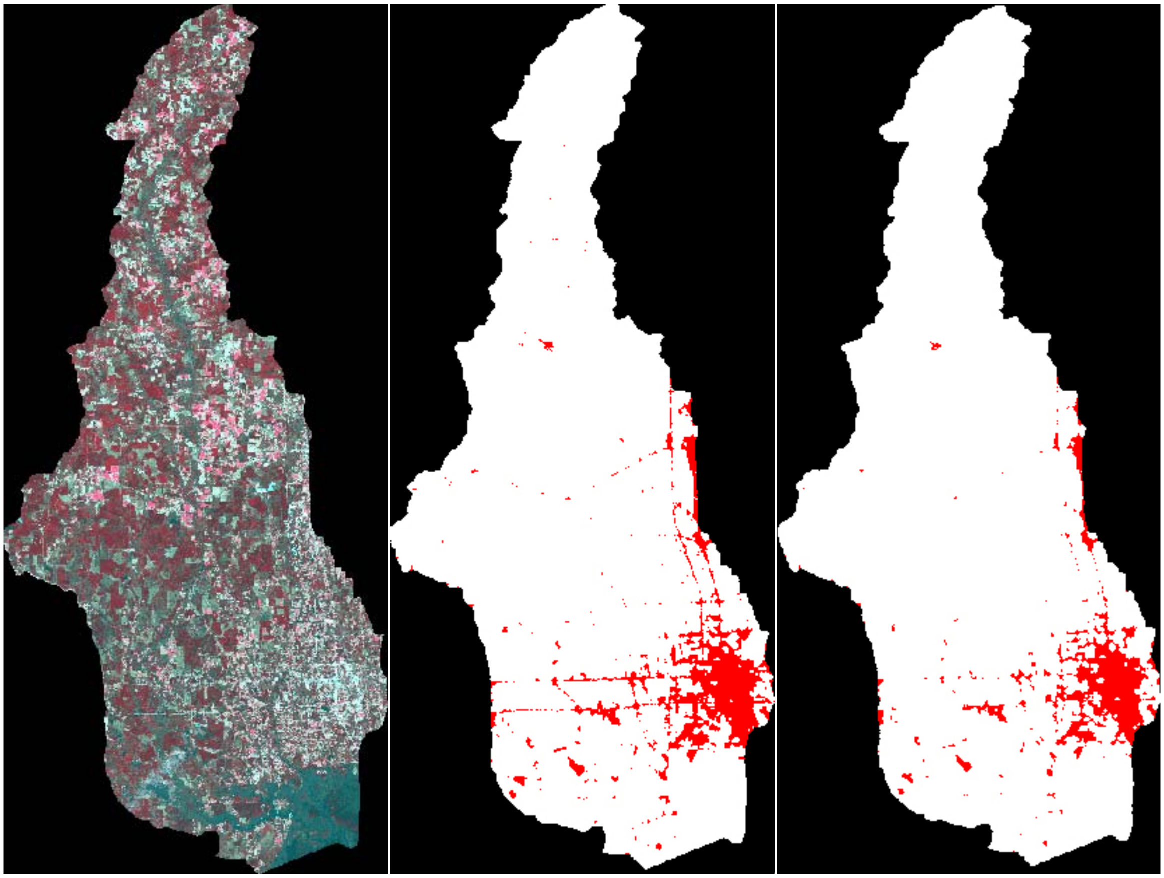
CA evolves at a sequential discrete time intervals or time steps. At each step, the state of cells will be updated simultaneously based on the transition rules. The classical CA assumes that the time steps are the same for all cells and transition rules are applied simultaneously at every cell. However, this assumption is not always valid in the real world. Different cells or even the same cell may have different time steps. Although the problem has been recognized, most studies still use static time steps. In this experiment, two sets of time step were selected based on the data availability: 2 - 3 years and 4 - 5 years. In order to examine the impact of time steps on the performance of CA-based LCLU modeling, two sets of data were used. The first set of data has 2 - 3 years time steps, and the second set of data has 4 - 5 years time steps. The results indicate that the overall accuracy decreased from 86.60% to 83.20% and the overall kappa value decreased from 0.7307 to 0.6714 when the time steps increased from 2 - 3 to 4 - 5 (Table 8.1). Figure 8.2 compares 2002 classified image and modeled images derived from various CA time step configurations. Images F and H represent the results corresponding to time step 2 - 3 years and 4 - 5 years respectively.

8.2.5. Transition Rules

The key component of CA is transition rules. The transition rules are a set of conditions or functions that specify how the state of cell changes according to its historical and current state. The future state of a cell is controlled by the transition rules, state, neighborhood, and time step. In general, there are two types of transition rules: implicit transition rules and explicit transition rules. Usually, the implicit transition rules are represented by mathematical expressions (Li and Yeh, 2004), such as linear equations (Wu and Webster, 1998), logistic models (Wu, 2002), fuzzy sets (Li and Yeh, 2000), and neural networks (Li and Yeh, 2002). The explicit transition rules are expressed by natural language that can be easily understood by decision makers (Li and Yeh, 2004). In order to examine the impact of transition rules on the performance of CA-based LCLU modeling, two sets of transition rules were developed under different conditions. The optimized transition rules were created by using the configuration: generation 500, population 300, crossover 0.59, mutation 0.05, and elitism 0.04. The un-optimized transition rules were created by using the configuration: generation 50, population 50, crossover 0.09, mutation 0.01, and elitism 0.01. The pruning rate 2% is for both of them. The results indicate that the overall accuracy changes from 86.60% to 73.40% and overall kappa value changes from 0.7307 to 0.5167 when transition rules were changed from optimized to un-optimized (Table 8.1). Figure 8.2 compares the 2002 classified image and a model derived from various CA configurations. Images C and F represent the results corresponding to un-optimized transition rules and optimized transition rules respectively.

8.2.6. Validation

Both calibration and validation are very important for spatio-temporal modeling. Calibration is the iterative process in which the output of model and the real world is compared, and the parameters of model are adjusted in order to improve the agreement between the output of the model and the real world. Its goal is to estimate the value of parameters of a model and to find a set of best-fit parameter values so that the model can efficiently simulate the real world. In fact, the above five tests on individual CA components have consisted of the process of calibration. Through the phase of calibration, the optimized CA components configuration can be determined. Validation of the optimized CA-based LCLU predictive model was then conducted (Table 8.2, and Figure 8.3). Validation is the process of examining how well the model's output characterizes the target system. Usually, it is done by comparing the output of the optimized predictive model and the classified image in the study area. The data used in the phase of calibration included Landsat TM / ETM images in 1990, 1992, 1995, 1997, 2000, and 2002. The data used in the phase of validation should be different from the data used in the phase of calibration. Since the data resource was limited in this research, the data used in the phase of validation was only slightly different. It included classified Landsat TM / ETM images in 1990, 1992, 1995, 1997, 2000, 2002, and 2005. Based on the calibration tests, the optimized CA components configuration was: space – 6025 x 12993 cells, state – 2 classes, neighborhood –



Original Image AOI | 2005 Classified Urban | 2005 Modeling Urban

Figure 8.3. Validating the optimized CA-based land-cover and land-use predictive model by comparing 2005 classified urban map and modeling urban map

Table 8.2. Validation results of cellular automata-based land-cover and land-use predictive modeling*

Class Name	**Tested Total**	**Number Correct**	**Producer Accuracy**	**User Accuracy**	**Overall Accuracy**	**Kappa Value**
Urban	200	145	72.59%	100.00%	89.00%	0.7598
Non-urban	300	355	100.00%	84.51%		

* Validation using the following calibrated parameters:
Space -- 6225 x 12993 cells
States -- 2 classes
Time steps -- 2-3 years
Neighborhood -- von Neumann 3 x 3
Transition rules -- optimized under condition of generation 500, population 300, crossover 0.59, mutation 0.05, elitism rate 0.04, and pruning rate 2%.

von Neumann 3 x 3, time step – 2 to 3 years, and transition rules – optimized. After predicting the LCLU in 2005 using the data obtained during 1990 – 2002, the predicted 2005 LCLU was compared with the classified 2005 image (Figure 8.3). The accuracy of predictive modeling was performed using classification accuracy assessment techniques and the 2005 classified image was used as the reference data. The accuracy of modeling was 89.00% (Table 8.2). After validating the model, it was used to predict future LCLU in 2007-2008, 2010, 2012-2013, and 2015 (Figure 8.4). In Figure 8.4, the small-circled areas show some differences between the predicted year and its previous time period, and the large-circled areas show major differences during 2005-2015.

8.3. Discussion

8.3.1. Impact of Space on Land-Cover Land-Use Predictive Modeling

The space of CA usually depends on the data. The study area can be represented by a different number of cells or pixels. When the number of cells is large or the spatial resolution of the classified image is high, the CA has a large space. In the test, the following sets of space were examined: 378 x 813, 1507 x 3249, and 6025 x 12993. After performing the spatio-temporal modeling as described in previous chapters, the resultant predicted LCLU images were created. In this case, only one LCLU feature, namely URBAN in 2002, was predicted for calibration purposes. The prediction accuracy assessment was done by comparing the predicted image and classified image in 2002. The results indicate that both large and small space can improve CA-based model in terms of prediction accuracy. Small space may be less complex and provide less information, whereas large space may be more complex and provide more information. The accuracy of spatio-temporal modeling depends on many factors. The balance between complexity and information is just one of them. It may play an important role in this test. For example, when space is in the mid-sized range, the complexity is higher than small

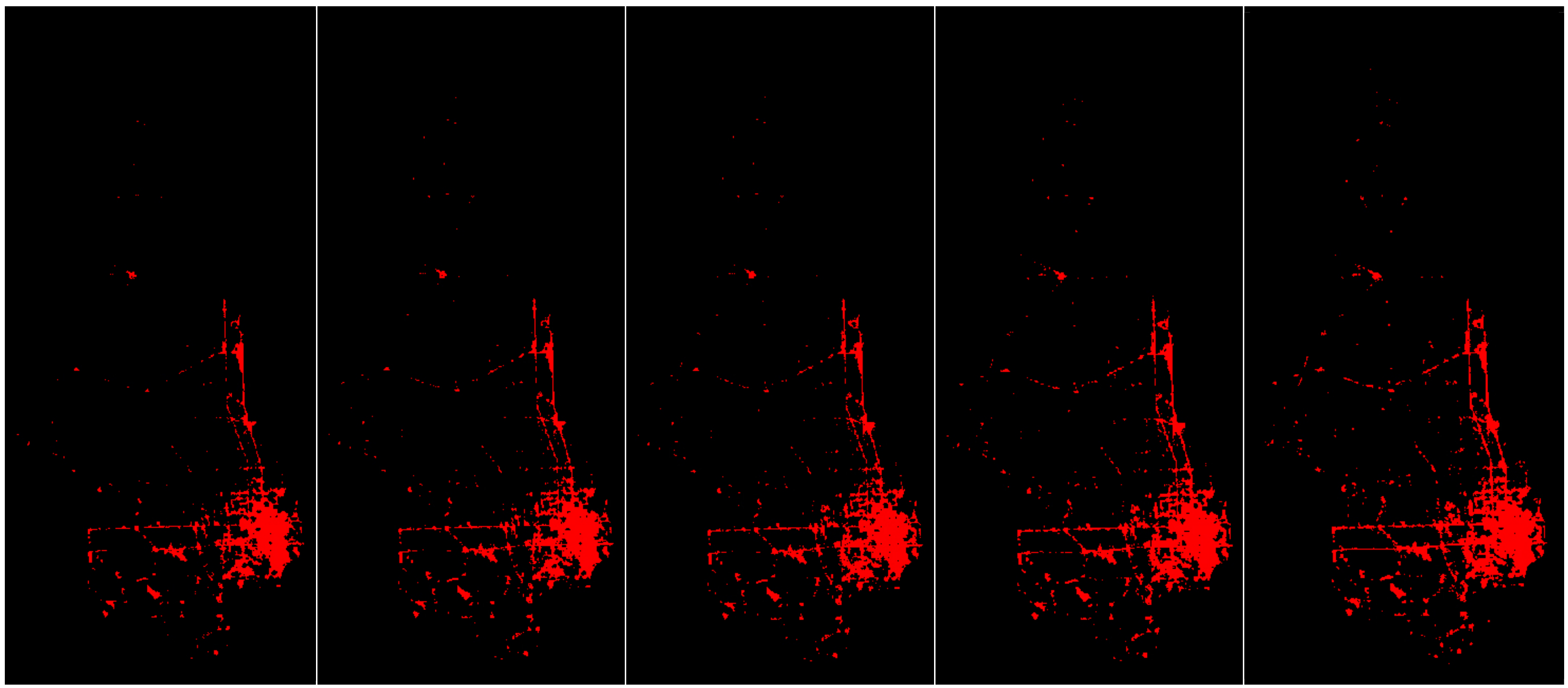


Figure 8.4. The prediction of urban area

space, and the information is less than large space, so the accuracy of spatio-temporal modeling is lower.

8.3.2. Impact of States on Land-Cover Land-Use Predictive Modeling

The number of states of a cell is determined based on the purpose of the modeling application. A basic CA model only requires a Boolean cell state, 0 or 1. Other models may have more than two states. As it has been mentioned in previous sections, two sets of states were selected for this test. The predicted URBAN in 2002 based on the two states of CA was different from the predicted URBAN in 2002 based on the seven states of CA. The CA with 2 states has a higher accuracy than the CA with 7 states (Table 8.1). A possible reason may be that the different features have a different influence on URBAN features. Furthermore, after converting the seven states into two states, the data were simplified, and the complexity of the data set decreases. When the process mining system has less information, it may make mistakes. On the other hand, the simplified data may make the development of the process model less prone to error. Therefore, there is a balance between simplicity and complexity. The impact of number of states on the accuracy of CA-based LCLU modeling is determined by the balance. In this case, simplicity played a crucial role. This may be the reason that a CA-based model with two states had higher accuracy.

8.3.3. Impact of Neighborhood on Land-Cover Land-Use Predictive Modeling

The type and size of neighborhood are important factors that affect the performance of any CA-based model. Although various neighborhoods have been used in geographic automata systems (Batty, 1998; Wu, 1998; White and Engelen, 1993, 1997; Li and Yeh, 2000; Yeh and Li, 2000, 2001, and 2002), no particular validation on what is the appropriate neighborhood type and size has been made in these applications. The two types of neighborhood tested were Moore and von Neumann. There were two sizes for each of them: 3 x 3 and 7 x 7. The modeling results indicate that the von Neumann neighborhood is better than Moore neighborhood, and a 3 x 3 neighborhood is better than a 7 x 7 neighborhood in terms of the accuracy of modeling (Table 8.1). Different shapes and sizes of neighborhoods have different capabilities to capture spatio-temporal information with different quality and quantity. Moore neighborhood is a square shape and von Neumann neighborhood is a rhombus shape. The large neighborhoods usually capture more information. With the same size, for example 3 x 3, a Moore neighborhood may capture more information than a von Neumann neighborhood because of the shape. However, the quality of information is a different and complex issue. According to Kocabas and Dragicevic (2006), a large neighborhood size can significantly reduce the predicted urban area compared to a small neighborhood size. The dissertation research results confirm that both large neighborhood size and Moore neighborhood reduce the predicted urban area more significantly than that of a small von Neumann neighborhood. The essential reason can be the shape and the size of neighborhood. With the same size, the rhombus shape has less coverage than the square shape. Another reason is that the shape of rhombus may fit the LCLU pattern better than the shape of square.

8.3.4. Impact of Time Steps on Land-Cover Land-Use Predictive Modeling

Time step is an important factor in spatio-temporal modeling. It includes at least two aspects: the number of steps and the length of each step. Based on the literature review, a fine time step can be days and months, and a coarse time step can be a year or years. Models with a fine time step, such as general ecosystem model (GEM) (Fitz *et al*., 1996), patuxent landscape model (PLM) (Voinov *et al*., 1999a, 1999b), and conversion of land use and its effects (CLUE) (Veldkamp and Fresco, 1996a and 1996b), can represent LCLU accurately (Agarwal *et al*., 2002). Models with coarse time step, such as land use change analysis system (LUCAS) (Berry *et al*., 1996), forest and agriculture sector optimization model (FASOM) (Adams *et al*., 1996), and land transformation model (LTM) (Pijanowski *et al*., 1997), may represent long-term LCLU changes better. Models with both fine and coarse time steps represent the temporal complexity better, but it is difficult to define them. The classic CA used the same time step for all cells and applies transition rules to every cell simultaneously (Singh, 2003). Despite having some limitations, time step-based CA, such as SLEUTH (slope, land use, exclusion, urban, transition, hill shading) are commonly used in LCLU modeling (US EPA, 2000). In order to determine the best time step for the CA used in this research, two sets of time steps were compared. The first one has 2-3 years time step length and a total of 4 steps, and the second one has 4-5 years time length and a total of 3 steps. The results indicate that CA with a small time step performs better than CA with a large time step. The possible reason is that high temporal resolution data can provide more detailed information about LCLU change. When predicting LCLU change, the CA with small time step usually leads to higher accuracy.

8.3.5. Impact of Transition Rules on Land-Cover Land-Use Predictive Modeling

The transition rules used in this research were developed based on the result of process-oriented LCLU change detection. The genetic PN-based process mining software takes LCLU event log data as input, builds process models using a PN algorithm, and optimizes the process models using GA. When process models were discovered under the optimized configuration of genetic parameters, the process models will represent the processes of LCLU change better. Therefore, the transition rules, based on those process models, will specify the states of cells at each time step better. If a set of transition rules were developed based un-optimized process models, they will not represent all LCLU change processes very well. The CA based on such poorly-defined transition rules cannot be used to predict LCLU accurately.

8.3.6. Calibration and Validation of Land-Cover Land-Use Predictive Model

The goal of calibration is to identify the best configuration of parameters, and the goal of validation is to examine how well the model’s output characterizes the target system. If the parameters of the model support its performance very well, the output of the model will represent the target system very well. In this dissertation, the validation procedure was carried out by using an optimized CA to generate 2005 urban maps, then comparing the modeling 2005 urban map with the classified 2005 urban map. Although the configuration of CA components was optimized, the output was not sufficient. The overall modeling accuracy was only 89.00% (Table 8.2). This result may suggest that the CA components optimization is a complex process. Although "trial and error" is an approach to optimize the CA components, its search space is limited, and the combination of the best individuals may not be the best solution.

8.4. Hypothesis #5 Review

The third research question as presented in Chapter 1 was how do cellular automata components (such as space, states, neighborhood, time steps, and transition rules) impact the accuracy of LCLU predictive modeling. The third research hypothesis presented in Chapter 1 was stated as follows:

> **Ha:** Different space size, state size, time step, neighborhood type and size, or transition rules can increase / decrease the accuracy of predictive modeling, respectively.

Based on the experiment designed and performed in this chapter, these CA components can impact the performance of CA-based LCLU predictive modeling significantly. In the case of space size, the relationship between space size and modeling accuracy is complex. When space sizes increased from 378 x 813 to 1507 x 3249, and then to 6025 x 12993, the overall urban growth modeling accuracies decreased from 90.60% to 86.60% and then increased to 93.00%.

In the case of number of states, although a larger number of states theoretically provides more detail information about different LCLU features, it may not benefit LCLU modeling. There is a balance between simplicity and complexity. The impact of the number of states on the accuracy of CA-based LCLU modeling is determined by the balance. In this case, the CA-based

LCLU model performs better when the state is the simple. When the number of states increased from 2 to 7, the accuracies of LCLU modeling decreased from 94.00% to 86.60%.

In the case of time step, the results indicate that CA with a small time step performed better than CA with large time step. The possible reason is that high temporal resolution data can provide more detailed information about LCLU change.

In the case of neighborhood type and size, the results illustrate the complex relationships among the accuracy of modeling, neighborhood type, and neighborhood size. Both large neighborhood size and Moore neighborhood reduced the prediction accuracy of the urban area more significantly than a small neighborhood and von Neumann neighborhood did. One reason is that the rhombus shape has less coverage than the square shape. Another reason is that the shape of rhombus may fit the LCLU pattern better than the shape of square.

In the case of transition rules, the results suggest that the transition rules based on optimized process models will better specify the states of cells at each time step. If a set of transition rules was developed based on un-optimized process models, they will not represent all LCLU change processes very well. The CA that are based on such poorly-defined transition rules will not predict LCLU accurately.

Finally, the results of calibration and validation suggest that the CA component optimization is a complex process. After calibration, the recommended CA component configuration includes: space - 6225x12993 cells, state - 2 classes, time step - 2 to 3 years, neighborhood - von Neumann 3x3, and transition rules - optimized. The overall accuracy of prediction is 89.00% with a kappa value of 0.7598.

8.5. Summary

This chapter analyzed the impacts of CA components on the performance of CA-based LCLU predictive modeling. The CA-based models were calibrated using classified 1990-2002

Landsat TM/ETM images and validated using 1992-2005 Landsat TM/ETM images. The following points summarize the major results and conclusions in this chapter:

- The configuration of CA components can impact the performance of CA-based LCLU predictive modeling significantly. When CA components were set up improperly (*e.g.*, small space size and un-optimized transition rules), the accuracy of LCLU modeling will be low.
- The recommended CA components configuration is: space - 6025 x 12993 cells, state – 2 classes, neighborhood – von Neumann 3 x 3, time step – 2 to 3 years, and transition rules – optimized.
- The urban area in 2005 is 153718 pixels. Based on the prediction, it will increase about 40% after ten years. The predicted urban area in 2015 is 215844 pixels.
- The process of calibration and validation of CA-based LCLU model is complex. Although "trial and error" is an approach to optimize the CA components, the calibration and validation of CA-based model is still a challenge.

CHAPTER 9 SUMMARY AND CONCLUSION

9.1. Research Findings

LCLU is an important manifestation of human-nature interaction. LCLU study plays a crucial role in the process of improving social-natural relationship. It includes at least three tasks: classification, change detection, and predictive modeling. The goal of this research is to systematically examine the human-nature relationship in the Tickfaw watershed, the impact of genetic parameters on classification and change detection, and the impact of CA components on predictive modeling. In the previous chapters, five experiments have been performed. This section provides a summary of the findings.

9.1.1. Impact of Genetic Parameters on Land-Cover and Land-Use Classification

Genetic parameters include the number of generations, population size, crossover rate, mutation rate, and generation gap or elitism rate. The relationship between the configuration of genetic parameters and the performance of GA-based LCLU classification is very complex. When genetic parameters are set improperly, for example, the number of generations or the size of population is too small, or the crossover rate, mutation rate, or elitism rate is too low or too high, premature convergence, local optimization, or unstable population occurs. These problems will lead to poor LCLU classification. In order to improve GA-based LCLU classification, the configuration of genetic parameters should be set at a moderate level. The recommended configuration is: generation 2000-5000, population 1000, crossover rate 69% - 99%, mutation rate 0.1% - 0.5%, and generation gap 25% - 50%.

9.1.2. Impact of Image Parameters on Land-Cover and Land-Use Classification

Image parameters include spatial resolution, training / testing data size, and data combinations. The relationship between the configuration of image parameters and the performance of GA-based LCLU classification is very complex. When the training / testing size is too small, the spatial and spectral knowledge will be developed in a small space. This will lead to premature convergence or local optimization. Both premature convergence and local optimization will lead to poor LCLU classification. When spatial resolution is too small or too large, the variability of inter- / intra classes and the proportion of mixed pixel will change, which will affect the performance of GA. The recommended configuration of image features and ancillary data is: 16 - 20 layers data combination, 10000 / 5000 – 20000 / 10000 training / testing data size, 30 m – 60 m spatial resolution, and optimum genetic parameter setting.

9.1.3. Comparison of GA-based and Traditional Land-Cover and Land-Use Classification

In this research, three traditional approaches were examined: ISODATA, MLC, and the hybrid of ISODATA and MLC. They were compared with GA-based approach in terms of the characteristics of algorithms, the performance of algorithms, and the operation of the algorithms. The characteristics of different approaches were stated in terms of algorithm type, assumptions, advantages, and disadvantages. The GA-based approach is a non-parametric approach, and it has no assumption on the distribution of data. The traditional approaches are statistics-based approaches, and they have assumption about the distribution of data (normal distribution). This assumption will limit the application of traditional approaches. The operations of all approaches were compared in terms of training data, the classification procedure, and time consumption. ISODATA is an approach that needs no training data and little time. Other approaches require training data, and GA-based approach requires much time. Finally, the performance of all approaches was evaluated in terms of classification accuracies of the overall level and individual level. At the overall level, the GA-based approach had the highest overall accuracy and kappa value; at the individual level, the GA-based approach had the highest producer accuracy, user accuracy, and kappa value in most land features classification. Based on the above information, it can be concluded that the GA–based approach is better than traditional approaches.

9.1.4. Impact of Genetic Parameters on Land-Cover and Land-Use Change Detection

The primary goal of this study was to investigate the relationship between genetic parameters and genetic PN-based LCLU change detection. The LCLU change process mining and accuracy was also examined. It was found that the relationships between genetic parameters and the performance of genetic PN-based LCLU change detection were very complex, and accuracy assessment of the LCLU process models is still a great challenge. GA parameters can impact the performance of GA-Petri net-based LCLU change detection significantly, especially the LCLU change process mining. When genetic parameters are set improperly, the average fitness of population, the best fitness of individual, and the coverage of process instances will be low. In order to improve the performance of genetic PN-based LCLU change detection, the number of generations and the size of population should be increased; crossover rate, mutation rate, and elitism rate should be kept at a middle level. The study has demonstrated the great advantage of genetic PN-based LCLU change detection. It can provide information about what, when, why, how, and how much LCLU changes. The process models can be used as transition rules in CA-based spatio-temporal modeling.

9.1.5. Impact of CA Components on Land-Cover and Land-Use Predictive Modeling

The components of CA include space, state, neighborhood type and size, time step length and step number, and transition rules. The configuration of CA components can impact the performance of CA-based LCLU predictive modeling significantly. When CA components are set up improperly (e.g., small space size, large time steps, and un-optimized transition rules), the accuracy of CA-based LCLU modeling will be low. In order to improve the performance of CA-based LCLU modeling, the recommended CA component configuration is: space -- 6025 x 12993 cells, state -- 2 classes, neighborhood – von Neumann 3 x 3, time step – 2 to 3 years, and transition rules – optimized.

9.2. Research Contributions

9.2.1. Theoretical Contribution

The research seeks to make the connections among LCLU classification, change detection, and predictive modeling by using the Tickfaw watershed LCLU changes as an example. The explicit and systematic spatio-temporal knowledge developed in the processes-oriented LCLU change detection will definitely improve our understanding of LCLU change. The study also improves our understanding on GA and CA. Although GA was inspired by the theory of genetics and natural selection, they do not automatically inherit the characteristics of global optimization from both. It is important to recognize that improper genetic parameters configuration can cause premature convergence and local optimization, hence inaccurate classification and change detection. Although CA were not necessarily inspired by the theory of cytology, they have been broadly used to simulate cell-based phenomena. Although CA have been used in geographic research for at least two decades, study on the impact of CA components on the performance of CA-based spatio-temporal modeling have seldom been reported. This dissertation research provides a systematical investigation on this issue. The results indicate that the component of CA heavily impact its performance, and that improper CA components configuration can cause poor LCLU predictive modeling.

9.2.2. Methodological Contribution

LCLU change detection is one of the most important tasks in LCLU study. The literature review has indicated that there is no proper approach to perform a process-oriented LCLU change detection in terms of efficiency, capability, and explicit spatio-temporal knowledge. The genetic PN not only provides a very promising way to perform such study, but also provides an efficient approach to developing transition rules for CA-based spatio-temporal modeling. The transition rule is the most critical component in CA. Its definition heavily depends on the domain knowledge and individual preferences. The challenge in CA-based modeling is how to obtain domain knowledge. The processes derived from genetic PN not only demonstrate detailed domain knowledge, but also can be directly translated into transition rules. Spatio-temporal modeling is an important aspect of geographic research that can be used to solve real world problems. CA is a major approach in spatiotemporal modeling. After introducing PN into the procedure of transition rules development, CA applications will be performed more easily in various research areas.

9.2.3. Practical Contribution

The Tickfaw watershed lies in the fastest developing area of Louisiana. It has experienced significant urban growth and forest segmentation during the last twenty years. The human-nature relationship has become one of the most important problem that impact the regional sustainable development. Although there have been several studies about this watershed (Wu, 2005; Couvillion, 2005; Lake Pontchartrain Basin Foundation, 2005; Demcheck et al., 2004), the study on LCLU change processes and future LCLU has not been reported yet. This dissertation research provides historic, current, and future LCLU maps that can be used by local government for assessing the regional environment and developing sustainable development plan.

9.3. Future Research

Several further research can be recommended:

- The evaluation of PN-based LCLU change process modeling – theoretically the process models with high fitness value can represent real LCLU change processes well, but practically this is not always true. More studies are needed to explore why this situation occurs and what factors are.
- The present study was carried out based on the limited resources. In order to

generalize the results and conclusion, more LCLU features, multiple scale study area, and various resolution of spatio-temporal data should be involved.

- Calibration, validation, and sensitivity analysis are very important for CA-based LCLU predictive modeling. Only calibration and validation were performed in this research. Sensitivity analysis should be included in the future work.
- The comparison of various LCLU classification approaches was based on ISODATA, MLC, hybrid of ISODATA and MLC, and optimized GA-based approach. In the future study, all classification approaches should be optimized before the comparison.
- Currently, most popular spatial analysis software cannot support spatio-temporal analysis and modeling. Integrating both spatial and temporal aspects into the same software will make geo-computation more efficient and more significant.

REFERENCES

Adams, D. M., Alig, R. J., Caollaway, J. M., McCarl, B. A., and Winnett, S. M. 1996. *The Forest and Agricultural Sector Optimization Model (FASOM): Model Structure and Policy Application. Res. PNW-RP-495*. Forest Service, Pacific Northwest Research Station, U.S. Department of Agriculture. Portland, OR. 1-68.

Adams, J. B., Sabol, D., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O., and Gillespie, A. R. 1995. Classification of Multitemporal Images based on Fractions of Endmembers: Application to Land-Cover Change in the Brazilian Amazon. *Remote Sensing of Environment*, 52(2):137-154.

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P, and Schweik, C. M. 2002. *A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. In: Gen. Tech. Rep. NE-297.* Forest Service, Northeastern Research Station, U.S. Department of Agriculture, Newton Square, PA. 1-61.

Alcamo, J., Leemans, R., and Kreileman, E. 1998. *Global Change Scenarios of the 21st Century*. Pergamon, Oxford. 1-308.

Alves De Medeiros, A. K., Weijters, A. J. M. M., and van der Aalst, W. M. P. 2007. Genetic Process Mining: An Experimental Evaluation. *Journal of Data Mining and Knowledge Discovery*, 14(2):245-304.

Anas, A. 1994. METROSIM: A Unified Economic Model of Transportation and Land-Use. Alex Anas & Associates, Williamsville, NY.

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data. USGS Professional Paper 964*. US Government Printing Office, Washington, D.C. 1-41.

Argialas, D. and Harlow, C. 1990. Computational Image Interpretation Models: An Overview and Perspective. *Photogrammetric Engineering and Remote Sensing*, 56(6):871-886.

Armstrong, M. P. and Bennett, D. A. 1990. A Bit-Mapped Classifier for Ground Water Quality Assessment. *Computer and Geosciences*, 16(6):811-832.

Armstrong, M. P., Xiao, N, and Bennett, D. A. 2003. Using Genetic Algorithms to Create Multicriteria Class Intervals for Choropleth Maps. *Annals of the Association of American Geographers*, 93(3):595-623.

Atkinson, P. M. and Curran, P. J. 1997. Choosing an Appropriate Spatial Resolution for Remote Sensing Investigations. *Photogrammetric Engineering and Remote Sensing*, 63(12):1345-1351.

Atkinson, P. M. and Aplin, P. 2004. Spatial Variation in Land Cover and Choice of Spatial Resolution for Remote Sensing. *International Journal of Remote Sensing*, 25(18):3687-3702.

Balzter, H., Braun, P. W., and Kohler, W. 1998. Cellular Automata Models for Vegetation Dynamics. *Ecological Modeling*, 107(2-3):113-125.

Bandyopadhyay, S. and Pal, S. K. 2001. Pixel Classification Using Variable String Genetic Algorithms with Chromosome Differentiation. *IEEE Transaction on Geoscience and Remote Sensing*, 39(2):303-308.

Bandyopadhyay, S. 2005. Satellite Image Classification Using Genetically Guided Fuzzy Clustering with Spatial Information. *International Journal of Remote Sensing*, 26(3):579-593.

Batty, M. 1991. Generating Urban Forms from Diffusion Growth. *Environment and Planning A*. 23(4):511-544.

Batty, M. 1998. Urban Evolution on the Desktop: Simulation with the Use of Extended Cellular Automata. *Environment and Planning A*, 30(11):1943-1967.

Batty, M. 2000. Geocomputation Using Cellular Automata. In: *Geocomputation*. Edited by: Openshaw, S., and Abrahart, B. Taylor and Francis, London. 95-126.

Batty, M. 2002. Thinking about Cities as Spatial Events. *Environment and Planning B*, 29(1):1-2.

Batty, M., Longley, P., and Fotheringham, S. 1989. Urban Growth and Form: Scaling, Fractal Geometry, and Diffusion-Limited Aggregation. *Environment and Planning A*, 21(11):1447-1472.

Batty, M. and Xie, Y. 1994. From Cell to Cities. *Environmental and Planning B*, 21(7):s31-s48.

Batty, M., and Xie, Y. 1997. Possible Urban Automata. *Environment and Planning B*, 24(2):175-192.

Batty, M., Xie, Y. C., and Sun, Z., L. 1999. Modeling Urban Dynamics Through GIS-based Cellular Automata. *Computers, Environment and Urban Systems*, 23 (31):205-233.

Bays, C. 1990. The Discovery of a New Glider in the Game of Three-Dimensional Life. *Complex Systems*, 4(6):599-602.

Bays, C. 1991. A New Game of Three-Dimensional Life. *Complex Systems*, 5(1):15-18.

Beauchemin, C., Samuel, J., and Tuszynski, J. 2005. A Simple Cellular Automata Model for Influenza A Viral Infections. *Journal of Theoretical Biology*, 232(2):223-234.

BenDor, T. and Metcalf, S. S. 2006. The Spatial Dynamics of Invasive Species Spread. *System Dynamics Review*, 22(1):27-50.

Bennett, D. A., Xiao, N. C., and Armstrong, M. P. 2004. Exploring the Geographic Consequences of Public Policies Using Evolutionary Algorithms. *Annual of Association of Geographers*, 94(4):827-847.

Berjak, S. G., and Hearne, J. W. 2002. An Improved Cellular Automata Model for Simulating Fire in a Spatially Heterogeneous Savanna System. *Ecological Modeling*, 148(2):133-151.

Berry, M. W., Hazen, B. C., Maclntyre, R. L., and Flamm, R. O. 1996. LUCAS: A System for Modeling Land-Use Change. *IEEE Computational Science and Engineering*, 3(1):24.

Bolstad, P. V. and Lillesand, T. M. 1992. Rule-based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data. *Photogrammetric Engineering and Remote Sensing*, 58(7):965-971.

Bond, J. E., Beamer, D. A., and Hedin, M. 2006. Combining Genetic and Geospatial Analysis to Infer Population Extinction in Mygalomorph Spiders Endemic to the Los Angeles Region. *Animal Conservation*, 9(2):145-157.

Braimoh, A. K. and Vlek, P. L. G. 2005. Land-Cover Change Trajectories in Northern Ghana. *Environmental Management*, 36(3):356-373.

Brookes, C. J. 2001. A Genetic Algorithm for Designing Optimal Patch Configurations in GIS. *International Journal of Geographical Information Science*, 15(6):539-559.

Brown, D.G., Walker, R., Manson, S., and Seto, K. 2004. Modeling Land Use and Land Cover Change. In: *Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface*. Edited by: Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Turner II, B. L., and Cochrane, M. A. Springer-Verlag, New York, NY. 395-409.

Brumby, S. P., Theiler, J., Perkins, S. J., Harvey, N., Szymanski, J. J., Bloch, J. J., and Mitchell, M. 1999. Investigating of Image Feature Extraction by a Genetic Algorithm. *Proc. SPIE*, 3812:24-31.

Brumby, S. P., Harvey, N. R., Perkins, S. J., Porter, R. B., Szymanski, J. J., Theiler, J. P., and Bloch, J. J. 2000. Genetic Algorithm for Combining New and Existing Image Processing Tools for Multispectral Imagery. *Proc. SPIE*, 4049:480-490.

Bulitko, V. and Wilkins, D. 2003. Qualitative Simulation of Temporal Concurrent Processes Using Time Interval Petri Nets. *Artificial Intelligence*, 144(1-2):95-124.

CADGIS (Computer Aided Design and Geographic Information System Research Laboratory at Louisiana State University). 1984. *Atlas: The Louisiana Statewide GIS*. http://www.atlas.lsu.edu. Last accessed on December 12, 2006.

Campbell, J. B. 1996. *Introduction to Remote Sensing* (2nd). The Guilford Press, New York, NY. 1-606.

Candau, J., Rasmussen, S. and Clarke, K. C. 2000. A Coupled Cellular Automata Model for Land Use/Land Cover Dynamics. In: *GIS/EM4 Proceedings*. 4th International Conference on Integrating GIS and Environmental Modeling: Problems, Prospects and Research Needs. Septmber 2-8, 2000. Banff, Alberta, Canada.

Candau, J. 2002. *Temporal Calibration Sensitivity of the SLEUTH Urban Growth Model (Thesis).* Department of Geography, University of California, Santa Barbara, CA. 1-110.

Carey, P. D. 1996. Disperse: A Cellular Automata for Predicting the Distribution of Species in a Changed Climate. *Global Ecology and Biogeography Letters*, 5(4/5):217-226.

Chalifoux, S., Cavayas, F. and Gray, J. T. 1998. Mapping-Guided Approach for the Automatic Detection on Landsat TM Images of Forest Stands Damaged by the Spruce Budworm. *Photogrammetric Engineering and Remote Sensing*, 64(6):629-635.

Chambers, L. 2000. *The Practical Handbook of Genetic Algorithms Application*. 2nd Edition. Chapman & Hall/CRC, New York, NY. 99-199.

Chalmers, N. and Fabricius, C. 2007. Expert and Generalist Local Knowledge about Land-Cover Change on South Africa's Wild Coast: Can Local Ecological Knowledge Add Value to Science? *Ecology and Society*, 12(1):10.

Chan, J. C.-W., Chan, K.-P., and Yeh, A. G.-O. 2001. Detecting the Nature of Change in an Urban Environment: a Comparison of Machine Learning Algorithms. *Photogrammetric Engineering & Remote Sensing*, 67(2):213-225.

Chang, C. K., Chang, Y. F., Song, C. C., and Aoyama, M. 1989, Integral: Petri-Net Approach to Distributed Software Development. *Information and Software Technology*, 31(10):535-545.

Chen, D., Stow, D. A. and Gong, P. 2004, Examining the Effect of Spatial Resolution and Texture Window Size on Classification Accuracy: An Urban Environment Case. *International Journal of Remote Sensing*, 25(11):2177-2192.

Chen, J. and Gong, P., He, C. Y., Luo, W., Masayuki, T., and Shi, P. J. 2002. Assessment of Urban Development Plan of Beijing by Using a CA-based Urban Growth Model. *Photogrammetric Engineering & Remote Sensing*, 68(10):1063-1071.

Chen, K. S., Tzeng, Y. C., Chen, C. F., and Kao, W. L. 1995. Land Cover Classification of Multispectral Imagery Using a Dynamic Learning Neural Network. *Photogrammetric Engineering and Remote Sensing*, 61(4):403-408.

Chen, L. 2003. A Study of Applying Genetic Programming to Reservoir Trophic State Evaluation Using Remote Sensor Data. *International Journal of Remote Sensing*, 24(11):2265-2275.

Cheng, G. L., Wang, X. F., Zhuang, Z. Q., and Wang, D. S. 1996. *Genetic Algorithms and Application*. People's Post and Telecommunication Press, Beijing, China. 1-433.

Chien, C. F., and Chen, C. H. 2007. Using Genetic Algorithms (GA) and a Colored Timed Petri Net (CTPN) for Modeling the Optimization-based Schedule Generator of a Generic Production Scheduling System. *International Journal of Production Research*, 45(8):1763-1789.

Civco, D. L., Hurd, J. D., Wilson, E. H., Song, M. J., and Zhang, Z. K. 2002. A Comparison of Land Use and Land Cover Change Detection Methods. *ASPRS-ACSM*. Annual Conference and FIG XXII Congress. April 22-26, 2002. Washington, D.C.

Clark, C., and Canas, A. 1995. Spectral Identification by Artificial Neural Network and Genetic Algorithm. *International Journal of Remote Sensing*, 16(12):2255-2275.

Clarke, K. C., Brass, J. A., and Riggan, P. J. 1994. A Cellular Automata Model of Wildfire Propagation and Extinction. *Photogrammetric Engineering & Remote Sensing*, 60:1355-1367.

Clarke, K. C., Hoppen, S., and Gaydos, L. J. 1997. A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area. *Environment and Planning B*, 24(2):247-261.

Clarke, K. C. and Gaydos, L. J. 1998. Loose-Coupling a Cellular Automata Model and GIS: Long-Term Urban Growth Prediction for San Francisco and Washington / Baltimore. *International Journal of Geographical Information Science*, 12(7):699-714.

Collins, J. B., and Woodcock, C. K. 1996. An Assessment of Several Linear Change Detection Techniques for mapping Forest Mortality Using Multitemporal Landsat TM Data. *Remote Sensing of Environment*, 56(1):66-77.

Congalton, R. G. 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37(1):35-46.

Cook, J. E. and Wolf, A. L. 1998. Event-based Detection of Concurrency. In: *Proceedings of the 6th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. Foundations of Software Engineering. Lake Buena Vista, Florida. 35-45.

Cooley, R. E., Hobbs, M. H. W., and Pack, A. D. 1997. Genetic Algorithms and the Analysis of Spatially Referenced Data. *Applied Artificial Intelligence*, 11(2):151-170.

Coppin, P. R., and Bauer, M. E. 1996. Digital Change Detection in Forest Ecosystems with Remote Sensing Imagery. *Remote Sensing Reviews*, 13(3-4):207-234.

Coppin, P., Jonckheere, I., Nackaerts, K., and Muys, B. 2004. Digital Change Detection Methods in Ecosystem Monitoring: A Review. *International Journal of Remote Sensing*, 25(9):1565-1596.

Couclelis, H. 1985. Cellular Worlds: A Framework for Modeling Micro-Macro Dynamics. *Environment and Planning A*, 17(5):585-596.

Couclelis, H. 1987. Cellular Dynamics: How Individual Decisions Lead to Global Urban Change. *European Journal Operational Research*, 30(3):344-346.

Coucelis, H. 1988. Of Mice and Men: What Rodent Populations Can Teach Us about Complex Spatial Dynamics. *Environment and Planning A*, 20(1):99-109.

Couvillion, B. R. 2005. *Spatial Heterogeneity in Forested Landscapes: An Examination of Forest Fragmentation and Suburban Sprawl in the Florida Parishes of Louisiana (Thesis).* Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA. 1-136.

Cushnie, J. L. 1987. The Interactive Effect of Spatial Resolution and Degree of Internal Variability within Land-Cover Types on Classification Accuracies. *International Journal of Remote Sensing*, 8(1):15-29.

David, R. and Alla, H. 1992. *Petri Nets & Grafcet: Tools for Modeling Discrete Event Systems*. Prentice Hall, New York, NY. 1-339.

De Jong, K. A. and Sarma, J. 1993. Generation Gap Revisited. In: *Foundations of Genetic Algorithms 2*, Edited by: Whitley, L. D. Morgan Kaufmann, San Matco, CA. 19-28.

Dean, A. M. and Smith, G. M. 2003. An Evaluation of Per-Parcel Land Cover Mapping Using Maximum Likelihood Class Probabilities. *International Journal of Remote Sensing*, 24(14):2905-2920.

DeFries, R. S., Hansen, M. Townshend, J. R. G., and Sohlberg, R. 1998. Global Land Cover Classifications at 8km Spatial Resolution: the Use of Training Data Derived from Landsat Imagery in Decision Tree Classification. *International Journal of Remote Sensing*, 19(6):3141-3168.

DeFries, R. S., Chan, J. C.-W. 2000. Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data. *Remote Sensing of Environment*, 74(3):503-515.

Delahaye, D. 2001. Airspace Sectoring by Evolutionary Computation. In: *Spatial Evolutionary Modeling*. Edited by: Krzanowski, R. and Raper, J. Oxford University Press, Inc., New York, NY. 203-233.

Demcheck, D. K., Tollett, R. W., Mize, S. V, Skrobialowski, S. C., Fendick, R. B., Swarzenski, J, C. M., and Porter, S. 2004. *Water Quality in the Acadian-Ponchartrain Drainages, Louisiana and Mississippi, 1999-2001. USGS Circular 1232*. Reston, VA. 1-33.

Demetris S., and Vasilakos A. 2006. Comparison of Computational Intelligence Based Classification Techniques for Remotely Sensed Optical Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2305-2318.

Department of Computer Science, University of Aarhus. 2007. *Examples of Industrial Use of CP-nets*. http://www.daimi.Au.dk/. Last accessed on March 4, 2007.

Desel, J., Oberweis, A., Reisig, W., and Rozenberg, G. 1998. *Petri Nets and Business Process Management*. http://www.dagstuhl.de/Reports/. Last accessed on March 4, 2007.

Desel, J. 2005. Process Modeling Using Petri Nets. In: *Process-Aware Information Systems: Bridging People and Software through Process Technology*. Edited by: Dumas, M., van der Aalst, W. M. P., and Hofstede, A. John Wiley & Sons, Inc., Hoboken, NJ. 147-176.

Deutsch, A. and Dormann, S. 2005. *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Application, and Analysis*. Birkhauser, Boston. 1-319.

Dibble, C. & Densham, P. J. 1993. Generating Interesting Alternatives in GIS and SDSS Using Genetic Algorithms, In: *Proc. GIS/LIS'93.* American Congress on Surveying and Mapping, Bethesda, MD. 1:180-189.

Dibble, C., 2001, Beyond Data: Handling Spatial and Analytical Contexts with Genetics-Based Machine Learning. In: *Spatial Evolutionary Modeling*. Edited by: Krzanowski, R. and Raper. J. Oxford University Press, Inc., New York, NY. 127-140.

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., and Luck, A. T. 2006. *Geographic Information Science and Technology Body of Knowledge*. Association of American Geographers, Washington, D.C. 1-148.

Dijk, S. V., Thierens, D. and Berg, M. D., 2001, Designing Genetic Algorithms to Solve GIS Problems. In: *Spatial Evolutionary Modeling.* Edited by: Krzanowski, R. and Raper, J. Oxford University Press, Inc., New York, NY. 158-178.

Dijk, S. V., Thierens, D. and Berg, M. D., 2002, Using Genetic Algorithms for Solving Hard Problems in GIS. *GeoInformatica*, 6(4):381-413.

Doi, A., Drath, R., Nagasaki, M, and Miyano, S. 1999. Protein Dynamics Observations of Lambda-Phage by Hybrid Petri Net. *Genome Informatics*, 10:217-218.

Dolman, A. J. and Verhagen, A. 2003. Land Use and Global Environmental Change. In: *Global Environmental Change and Land Use*. Edited by: Dolman, A. J., Verhagen, A., and Rovers, C. A. Kluwer Academic Publishers, Boston, MA. 3-16.

Dos Santos, R. M. Z., and Coutinho, S. 2001. Dynamics of HIV Infection: A Cellular Automata Approach. *Physical Review Letters*, 87(16):1-4.

Dowling, R., Ireson, R., Skabardonis, A., Gillen, D., Stopher, P., Horowitz, A., Bowman, J., Deakin, E., and Dulla, R. 2000. *Predicting Short-Term and Long-Term Air Quality Effects of Traffic-Flow Improvement Projects. Technical Report NCHRP 25-21*. Transportation Research Board, National Research Council, Washington, D.C.

Dustdar, S., Hoffmann, T. and van der Aalst, W. M. P. 2005. Mining of ad-hoc Business Processes with TeamLog. *Data and Knowledge Engineering*, 55(2):129-158.

EROS (Earth Resources Observation and Science). 2005. *Satellite products Description*. http://edc.usgs.gov/products/satellite.html, Last Accessed on may 1, 2005.

Ehlers, M., Edwards, G., and Bedard, Y. 1989. Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution. *Photogrammetric Engineering and Remote Sensing*, 55(11):1619-1627.

Electric Power Research Institute (EPRI). 2002. *SmartPlaces Version 4.0.* http://www.epri.com. Last Accessed on March 4, 2007.

Engelen, G., White, R., and Uljee, I. 1997. Integrating Constrained Cellular Automata Models, GIS and Decision Support Tools for Urban Planning and Policy Making. In: *Decision Support Systems in Urban Planning*. Edited by: Timmermans, H. P. J. E&FN Spon, London. 125-155.

Ferscha, A. 1992. A Petri Net Approach for Performance Oriented Parallel Program Design. *Journal of Parallel and Distributed Computing*, 15(3):188-206.

Friedl, M. A., Brodley, C. E. and Strahler, A. H. 1999. Maximizing Land Cover Classification Accuracies by Decision Tree at Continental to Global Scales. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):969-977.

Fitz, H. C., DeBellevue, E. B., Cpstanza, R., Boumans, R., Maxwell, T., Wainger, L., and Sklar, F. H. 1996. Development of General Ecosystem Model for a Range of Scales and Ecosystems. *Ecological Modeling*, 88(1-3):263-295.

Foody, G. M., McCulloch, M. B., and Yates, W. B. 1995. Classification of Remotely Sensed Data by an Artificial Neural Network: Issues Related to Training Data Characteristics. *Photogrammetric Engineering and Remote Sensing*, 61(4):391-401.

Forster, B. C. 1984. Derivation of Atmospheric Correction Procedures for Landsat MSS with Particular Reference to Urban Data. *International Journal of Remote Sensing*, 5(5):799-817.

Freeman, S. 2002. *Biological Science*. Prentice Hall, Upper Saddle River, NJ. 186-692.

Freitas, A. A. 2002. *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer-Verlag, New York, NY. 1-263.

Friedl, M. A. and Brodley, C. E. 1997. Decision Tree Classification of Land Cover from Remotely Sensed Data. *Remote Sensing of Environment*, 61(3):399-409.

Friedl, M. A., Brodley, C. E. and Strahler, A. H. 1999. Maximizing Land Cover Classification Accuracies by Decision Tree at Continental to Global Scales. *IEEE Transactions on Geoscience andRemote Sensing*, 37(2):969-977.

Gao, B. 1996. NDWI – A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sensing Environment*, 58(3):257-266.

Gao, Y., Mas, J. F., Maathuis, B. H. P., Zhang, X. M., Van Dijk, P. M. 2006. Comparison of Pixel-based and Object-oriented Image Classification Approaches – a Case Study in a Coal Fire Area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing*, 27(18):4039-4055.

Garcia-Orellana, C. J., Macias-Macias, M., Serrano-Perez, A., Gonzalez-Velasco, H. M., Gallardo-Caballero, R. 2002. A Comparison of PCA, ICA, and GA Selected Feature for Cloud Field Classification. *Journal of Intelligent & Fuzzy Systems*, 12(3-4):213-219.

Gardner, M. 1970. The Fantastic Combination of John Conway's New Solitaire Game "life", *Scientific American*, 223(4):120-123.

Garzelli, A., and Nencini, F. 2006. PAN-Sharpening of Very High Resolution Multispectral Images Using Genetic Algorithms. *Journal of Intelligent & Fuzzy Systems*, 27(15):3273-3292.

Gaylord, D. J. and Mishidate, K. 1996. *Modeling Nature: Cellular Automata Simulations with Mathematics*. Springer-Verlag, New York, NY. 1-37.

Geneletti, D. and Gorte, B. G. H. 2003. A Method for Object-Oriented Land Cover Classification Combining Landsat TM Data and Aerial Photographs. *International Journal of Remote Sensing*, 24(6):1273-1286.

GLCF (Global Land Cover Facility). 2005. *Data and Products*. http://glcf.umiacs.umd.edu/. Last Accessed on June 6, 2005.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, New York, NY. 1-432.

Goodchild, M. F. 1992. Geographical Information Science. *International Journal of Geographical Information Systems*. 6(1):31-45.

Gou, H., Huang, B. Q., Liu, W., Ren, S., and Li, Y. 2000. Petri-Net-based Business Process Modeling for Virtual Enterprises. In: *SMC'2000 Proceedings*. IEEE International Conference on Systems, Man, and Cybernetics, October 8-11, 2000. Nashville, TN. 5:3183-3188

Gourgand, J. M. 1993. Petri Net based Methodology for Task Scheduling on Multiprocessor Architectures. *Simulation*, 61(3):185-192.

Green, D. G., House A. P., and House, S. M. 1985. Simulating Spatial Patterns in Forest Ecosystems. *Mathematical and Computers in Simulation*, 27(2-3):191-198.

Grefensette, J. J. 1986. Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*. 16(1):122-128.

Gronewold, A. and Sonnenschein, M. 1998. Event-based Modeling of Ecological Systems with Asynchronous Cellular Automata. *Ecological Modeling*, 108(1-3):37-52.

Guerra, F., Puig, H., and Chaume, R. 1998. The Forest-Savanna Dynamics from Multi-date Landsat-TM Data in Sierra Parima, Venezuela. *International Journal of Remote Sensing*, 19(11):2061-2075.

Haack, B., Bryant, N., and Adams, S. 1987. An Assessment of Landsat MSS and TM Data for

Urban and Near-Urban Land-Cover Digital Classification. *Remote Sensing of Environment*, 21(2):201-213.

Hagerstrand, T. 1967. *Innovation Diffusion as a Spatial Process*. University of Chicago Press, Chicago, IL. 1-334.

Hagerstrand, T. 1968. A Monte Carlo Approach to Diffusion. In: *Spatial Analysis: A Reader in Statistical Geography*, Edited by: Berry, J. L., and Marble, D. F., Prentice Hall, Englewood Cliffs, NJ. 234-254.

Halid, M. 1997. Land Use-Cover Change Detection Using Knowledge based Approaches: Remote Sensing and GIS. In: *ACRS'1997 Proceedings*. 18th Asian Conference on Remote Sensing, October 20-24, 1997. Malaysia.

Hall, F. G., Strebel, D. E., Nickeson, J. E., and Goetz, S. J. 1991. Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images. *Remote Sensing Environment*, 35(1):11-27.

Hamadi, R., and Benatallah, B. 2003. A Petri net-based model for Web Service Composition. In: *Proceedings of the 14th Australasian Database Conference*. 17:191-200. Adelaide, Australia.

Hame, T., Heiler, I., and Miguel-Ayanz, J. S. 1998. An Unsupervised Change Detection and Recognition System for Forestry. *International Journal of Remote Sensing*, 19(6):1079-1099.

Hansen, M., Dubayah, R., and DeFries, R. 1996. Classification Tree: an Alternative to Traditional Land Cover Classifiers. *International Journal of Remote Sensing*, 17(5):1075-1081.

Harada, Y., and Iwasa, Y. 2006. Lattice Population Dynamics for Plants with Dispersing Seeds and Vegetation Propagation. *Research on Population Ecology*, 36(2):237-249.

Harvey, N. R., Brumdy, S., Perkins, S., Porter, R. B., Theiler, J. P., Young, A. C., Szymanski, J. J., and Bloch, J. J. 2000, Parallel Evolution of Image Processing Tools for Multispectral Imagery. *Proc. SPIE*, 4132:72-82.

Harvey, N. R., Theiler, J., Brumby, S. P., Perkins, S., Szymanski, J. J., Bloch, J. J., Porter, R. B., Galassi, M., and Young, A. C. 2002. Comparison of GENIE and Conventional Supervised Classifiers for Multispectral Image Feature Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2):393-404.

Herbert, J. and Stevens, B. H. 1960. A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science*, 2(2):21-36.

Herbst, J. and Karagiannis, D. 2004. Workflow Mining with InWoLvE. *Computers in Industry*, 53(3):245-264.

Herold, M., Goldstein, N. C., and Clarke, K. C. 2003. The Spatiotemporal form of Urban Growth: Measurement, Analysis and Modeling. *Remote Sensing of Environment*, 86(3):286-302.

Hinton, J. C. 1996. GIS and Remote Sensing Integration for Environmental Applications. *International Journal of Geographic Information Systems*. 10(7):877-890.

Hirata, Y., Matsuno, H., Sasaki, M., and Miyano, S. 2001. Genomic Object Net: Petri Net Enhancement for Multi-Cellular Processes. *Genome Informatics*, 12:292-293.

Hodgson, M. E., Jensen, J. R., Jason, Tullis, J. A., Riordan, K. D., and Archer, C. M. 2003. Synergistic Use Lidar and Color Aerial Photography for Mapping Urban Parcel Imperviousness. *Photogrammetric Engineering and Remote Sensing*, 69(9):973-980.

Holland, J. H. 1975. *Adaptation in Natural and Artificial System*. University of Michigan Press, Ann Arbor, MI. 1-211.

Holloway, L. E., Krogh, B. H., and Giua, A. 1997. A Survey of Petri Net Methods for Controlled Discrete Event Systems. *Discrete Event Dynamic Systems*, 7(2):151-190.

Hosage, C. M. and Goodchild, M. G. 1986. Discrete Space Location-Allocation Solutions from Genetic Algorithms. *Annals of Operations Research*, 6(2):35-46.

Hsieh, S. and Chen, Y. F. 1999. AgvSimNet: A Petri-Net-Based AGVS Simulation System. *The International Journal of Advanced Manufacturing Technology*, 15(11):851-861.

Hsieh, P. F., Lee, L. C., and Chen, N. Y. 2001. Effect of Spatial Resolution on Classification Errors of Pure and Mixed Pixels in Remote Sensing. *IEEE Transaction on Geoscience and Remote Sensing*, 39(12):2657-2663.

Hsu, P. Y., Chang, Y. B., and Chen, Y. L. 2003. STRPN: A Petri-Net Approach for Modeling Spatial-Temporal Relations between Moving Multimedia Objects. *IEEE Transactions on Software Engineering*, 29(1):63-76.

Huang, B., Cheu, R. L., and Liew, Y. S. 2004. GIS and Genetic Algorithms for HAZMAT Route Planning with Security Considerations. *International Journal of Geographical Information Science*, 18(8):769-787.

Huang, X. 1996. *A Machine Learning Approach to Automata Construction of Knowledge Bases for Expert Systems Remote Sensing Image Analysis with GIS Data (Dissertation).* Department of Geography, University of South Carolina, Columbia, SC. 1-102.

Huang, X. and Jenson, J. R. 1997. A Machine-Learning Approach to Automated Knowledge-Base Building for Remote Sensing Image Analysis with GIS Data. *Photogrammetric Engineering & Remote Sensing*, 63(10):1185-1194.

Huguenin, R. L., Karaska, M. A., Van Blaricom, D., Jensen, J. R. 1997. Subpixel Classification of Bald Cypress and Tupelo Gum Trees in Thematic Mapper Imagery. *Photogrammetric Engineering andRemote Sensing*, 63(6):717-725.

Hunt, J. D. 1997. *A Description of the MEPLAN Framework for Land Use and Transportation Interaction Modeling*. The 77th Annual Meeting of the Transportation Research Board, January 11-15, 1997. National Research Council, Washington, D.C.

Hussin, Y. A., Gier, A. de, and Hargyono. 1994. Forest Cover Change Detection Analysis Using Remote Sensing: A Test for the Spatially Resolved Area Production Model. In: *EGIS'94 Proceedings*. Fifth European Conference and Exhibition on Geographic Information Systems, March 29-April 1, 1994. Paris, France. 2:1825-1834.

Hutchinson, C. R. 1982. The Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement. *Photogrammetric Engineering and Remote Sensing*, 48(1):123-130.

Irons, J. R., Markham, B. L., Nelson, R. F., Toll, D. L., Williams, D. L., Latty, R. S., and Stauffer, M. L. 1985. The Effects of Spatial Resolution on the Classification of Thematic Mapper Data. *International Journal of Remote Sensing*, 6(8):1385-1403.

Jaragh, M. and Saleh, K. 1999. Synthesis of Communication Protocol Converters Using the Time Petri Net Model. *Journal of Systems and Software*, 47(1):53-69, 17.

Jenerette, G. D. and Wu, J. 2004. Analysis and Simulation of Land-Use Change in the Central Arizona – Phoenix Region. *Landscape Ecology*, 16(7):611-626.

Jensen, J. R. 1996. Issues Involving the Creation of Digital Elevation Models and Terrain Corrected Otthoimagery Using Soft-Copy Photogrammetry. In: *Digital Photogrammetry: An Addendum to the Manual of Photogrammetry*. Edited by: Greve, C. American Society for Photogrammetry and Remote Sensing. Bethesda, MD. 167-179.

Jeon, B. K., Jang, J. H. and Hong, K. S. 2002. Road Detection in Spaceborne SAR Images Using a Genetic Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 40(1):22-29.

Jin, Y. -Q. and Wang, Y. 2001. A Genetic Algorithm to Simultaneously Retrieve Land Surface Roughness and Soil Wetness. *International Journal of Remote Sensing*, 22(16):3093-3099.

Johnson, D. C. and Yodis, E. G. 1998. *Geography of Louisiana*. McGraw-Hill. USA. 1-207.

Johnson, R. D. and Kasischke, E. S. 1998. Change Vector Analysis: a Technique for the Multitemporal Monitoring of Land Cover and Condition. *International Journal of Remote Sensing*, 19(3):411-426.

Johnston, R. A., Shabazian, D. R., and Gao, S. Y. 2003. UPLAN: A Versatile Urban Growth Model for Transportation Planning. *Transportation Research Record*, 1831:202-209.

Johnston, R. J. 1997. *Geography & Geographers: Anglo-American Human Geography Since 1945*. Oxford University Press Inc., New York, NY. 53-102.

Jones, R. 2005. A Review of Land Use/Land Cover and Agricultural Change Models. In: *PIER Energy-Related Environmental Research. Cec-500-2005-056.* Stratus Consulting Inc. for the California Energy Commission, CA. 1-38.

Joy, S. M., Reich, R. M., and Reynolds, R. T. 2003. A Non-Parametric, Supervised Classification of Vegetation Types on the Kaibab National Forest Using Decision Tree. *International Journal of Remote Sensing*, 24(9):1835-1852.

Kain, J. F. 1986. Computer Simulation Models of urban Location. In: *Handbook of Regional and Urban Economics.* Edited by: Mills, E. S. Amsterdam, North-Holland. 2:847-875.

Kalles, D. and Papagelis, A. 2006. Analyzing Student performance in Distance Learning with Genetic Algorithm and Decision Tree. *Applied Artificial Intelligence*, 20(8):655-674.

Kanoh, H. and Wu, Y. 2003. Evolutionary Design of Rule Changing Cellular Automata. In: *Lecture Notes in Computer Science*. Springer-Verlag, New York, NY. 2773:258-264.

Karafyllidis, I., and Thanailakis, A. 1997. A Model for Predicting Forest Fire Spreading Using Cellular Automata. *Ecological Modeling*, 99(16):87-97.

Kartikeyan, B., Gopalkrishna, B., Kalubarme, M. H., and Majumder, K. L. 1994. Contextual Techniques for Classification of High and Low Resolution Remote Sensing Data. *International Journal of Remote Sensing*, 15(5):1037-1051.

Khorram, S. 1999. *Accuracy Assessment of Remote Sensing-Driven Change Detection*. American Society for Photogrammetry and Remote Sensing, Bethesda, MD. 1-61.

Klosterman, R. E. 1999. The What If? Collaborative Planning Support System. *Environment and Planning B*, 26(3):393-408.

Kocabas, V., and Dragicevic, S. 2006. Assessing Cellular Automata Model behaviour Using a Sensitivity Analysis Approach. *Computers, Environment and Urban Systems*, 30(6):921-953.

Kontoes, C., Wilkinson, G. G., Burrill, A., Goffredo, S., Megier, J. 1993. An Experimental System for the Integration of GIS Data in Knowledge-based Image Analysis for Remote Sensing or Agriculture. *International Journal of Geographic Information Systems*, 7(3):247-262.

Krzanowski, R. and Raper, J. 2001. *Spatial Evolutionary Modeling*, Oxford University Press, Inc. New York, NY. 63-126.

Kulkarni, A. D. and Lulla, K. 1999. Fuzzy Neural Network Models for Supervised Classification: Multispectral Image Analysis. *Geocarto International*, 14(4):42-50.

Kumar, R., Kumar, S., and Tiwari, M. K. 2005. An Expert Enhanced Colored Fuzzy Petri Net Approach to Reconfigurable Manufacturing Systems Involving Information Delays. *The International Journal of Advanced Manufacturing Technology*, 26(7-8):922-933.

Kuroda, C. and Ogawa, K. 1994. Nonlinear Waves in a Shear-Flow with a Diffusive Exothermic Reaction and Its Qualitative Reasoning. *Chemical Engineering Science*, 49(16):2699-2708

Lake Pontchartrain Basin Foundation. 2005. *Comprehensive Habitat Management Plan for the Lake Pontchartrain Basin*. http://www.saveourlake.org/. Last accessed on March 4, 2006.

Lambin, E. F., and Strahler, A. H. 1994. Indicators of Land-Cover Change for Change Vector Analysis in Multitemporal Space at Coarse Spatial Scales. *International Journal of Remote Sensing*, 15(10):2099-2119.

Lambin E. F. 1996. Change Detection at Multiple Temporal Scales: Seasonal and Annual Variations in Landscape Variables. *Photogrammetric Engineering and Remote Sensing*. 62(8):931-938.

Lambin, E. F. 2004. Modeling Land-Use Change. In: *Environmental Modeling: Finding Simplicity in Complexity*. Edited by: Wainwright, J. and Mulligan, M. John Wiley & Sons, Ltd., San Francisco, CA. 225-255.

Landgrebe, D. A., Biehl, L. L., and Simmons, W. R. 1977. An Empirical Study of Scanner System Parameters. *IEEE Transactions on Geoscience Electronics*, GE-15:120-130.

Landis, J. D. 1994. The California Urban Futures Model: A New Generation of Metropolitan Simulation Models. *Environment and Planning B*, 21(4):399-420.

Landis, J. D. 1995. Imagining Land Use Futures: Applying the California Urban Futures Model, *APA Journal*, 61(4):438-457.

Landis, J. D., and Zhang, M. 1998a. The Second Generation of California urban Futures Model: Part 1: Model Logic and Theory. *Environment and Planning B,* 25(5):657-666.

Landis, J. D., and Zhang, M. 1998b. The Second Generation of California urban Futures Model: Part 2: Specification and Calibration Results of the Land-Use Change Sub-model. *Environment and Planning B*, 25(6):795-824.

Latty, R. S., Nelson, R., Markham, B., Williams, D., and Toll, D. 1985. Performance Comparison between Information Extraction Techniques Using Variable Spatial Resolution Data. *Photogrammetric Engineering and Remote Sensing*, 51(9):1459-1470.

Lawrence, R. L., and Ripple, W. J. 1999. Calculating Change Curves for Multitemporal Satellite Imagery: Mount St. Helens 1980-1995. *Remote Sensing of Environment*, 67(3):309-319.

Legates, D. R., Gopal, S. and Rogerson, P. 2003. Mathematical Models and Quantitative

Methods. In: *Geography in America at the Dawn of the 21st Century*. Edited by: Gaile, G. L., and Willmott, C. J. Oxford: Oxford University Press, New York, NY. 442-450.

Li, X., and Yeh, A. G. O. 2000. Modeling Sustainable Urban Development by the Integration of Constrained Cellular Automata and GIS. *International Journal Geographical Information Science*, 14(2):131-152.

Li, X., and Yeh, A. G. 2002. Neural-Network-Based Cellular Automata for Simulating Multiple Land Use Changes Using GIS. *International Journal of Geographical Information Science*, 16(4):323-343.

Li, X. and Yeh, A. G.-O. 2004. Data Mining of Cellular Automata's Transition Rules. *International Journal of Geographical Information Science*, 18(8):723-744.

Li, X. and Yeh, A. G.-O. 2005. Integration of Genetic Algorithms and GIS for Optimal Locations Search. *International Journal of Geographical Information Science*, 19(5):581-601.

Liu, J., Zixing, Cai, Z. and Liu, J. 2000. Premature Convergence in Genetic Algorithm: Analysis and Prevention Based on Chaos Operator. In: *Proceeding of the 3rd World Congress on Intelligent Control and Automation*, June 28-July 2, 2000. Hefei, China. 495-499.

Liu, X., and Lathrop, R. G. Jr. 2002. Urban Change Detection based on an Artificial Neural Network. *International Journal of Remote Sensing*, 23(120):1501-1518.

Liu, Y. S., and Hao, Z. X. 2005. Petri Nets based Spatial-Temporal Relationships for Moving Objects. *Journal of Computer Sciences*, 1(4):510-514.

Lo, C. P. and Yeung, K. W. 2007. *Concepts and Techniques in Geographic Information Systems*. Pearson Prentice Hall, Upper Saddle Rive, NJ. 1-496.

Louis, S. J. and Rawlins, G. J. E., 1993, Predicting Convergence Time for Genetic Algorithms, In: *Foundations of Genetic Algorithms 2*. Edited by: Whitley, L. D. Morgan Kaufmann, San Matco, CA. 141-161.

Loveland, T. R., Sohl, T., Sayler, K., Gallant, A., Dwyer, J., Vogelmann, J., Zylstra, G., Wade. T., Edmonds, C., Chaloud, D., Jones, B. 2000. Land Cover Trends: Rates, Causes, and Consequences of Late-Twentieth Century U.S. Land Cover Change. In: *EPA/600/R-99/105 (NTIS PB2001-100348)*. U.S. Environmental Protection Agency, Washington, D.C. 1-36.

Lu, D. S. 2001. *Estimation of Forest Stand Parameters and Application in Classification and Change Detection of Forest Cover Types in the Brazilian Amazon Basin (Dissertation).* India State University, Terre Haute, IN. 1-235.

Lu, D., Mausel, P., Brondizio, E., and Moran, E. 2004a, Change Detection Techniques. *International Journal of Remote Sensing*, 25(12):2365-2407.

Lu, W., Gopal, S., and Woodcock, C. E. 2004b. Uncertainty and Confidence in Land Cover Classification Using a Hybrid Classifier Approach. *Photogrammetric Engineering and Remote Sensing*, 70(8):963-972.

Lu, D. and Weng, Q. 2004. Spectral Mixing Analysis of the Urban Landscapes in Indianapolis with Landsat ETM+ Imagery. *Photogrammetric Engineering and Remote Sensing*, 70(9):1053-1062.

Lu, D. and Weng, Q. 2007. A Survey of Image Classification Methods and Techniques for Improving Classification Performance. *International Journal of Remote Sensing*, 28(5):823-870.

Lucieer, A. and Kraak, M. 2004. Interactive and Visual Fuzzy Classification or Remote Sensing Imagery for Exploration or Uncertainty. *International Journal of Geographic Information science*, 18(5):491-512.

Ludeke, A. K., Maggio, R. C., and Reid, L. M. 1990. An Analysis of Anthropogenic Deforestation Using Logistic Regression and GIS. *Journal of Environmental Management*. 31(3):247-259.

Lyon, J. G., Yuan, D., Lunetta, R. S., and Elvidge, C. D. 1998. A Change Detection Experiment Using Vegetation Indices. *Photogrammetric Engineering and Remote Sensing*, 64(2):143-150.

Macomber, S. A., and Woodcock, C. E. 1994. Mapping and Monitoring Conifer Mortality Using Remote Sensing in the Lake Tahoe Basin. *Remote Sensing of Environment*, 50(3):255-266.

Mackett, R. L. 1990a. The Systematic Application of the LILT Model to Dortmund, Leeds and Tokyo. *Transportation Reviews*, 10(4):323-338.

Mackett, R. L. 1990b. Comparative Analysis of Modeling Land-Use Transport Interaction at the Micro and Macro Levels. *Environment and Planning A*, 22(4):459-275.

Mann, W. 1995. *Land Use/Transportation Integrated Model*. The 74th Annual Meeting of the Transportation Research Board. January 22-28, 1995. National Research Council, Washington, D.C.

Mas, J.-F. 1999. Monitoring Land-Cover Changes: A Comparison of Change Detection Techniques. *International Journal of Remote Sensing*, 20(1):139-152.

Matthews, K. B., Craw, S., Elder, S., Sibbald, A., and Mackenzie, I. 2000, Applying Genetic Algorithms to Multi-Objective Land Use Planning. In: GECCO'2000 *Proceedings.* Genetic and Evolutionary Computation Conference, July 8-12, 2000. Las Vegas, Nevada. 613-620.

Maulik, U. and Bandyopadhyay, S. 2003. Fuzzy Partitioning Using a Real Coded variable-Length Genetic Algorithm for Pixel Classification. *IEEE Transactions on Geoscience* *and Remote Sensing*, 41(5):1075-1081.

McCauley, S. and Goetz, S. J. 2004. Mapping Residential Density Patterns Using Multi-temporal Landsat Data and Decision Tree Classifier. *International Journal of Remote Sensing*, 25(6):1077-1094.

Mennis, J. and Liu, J. W. 2005. Mining Association Rules in Spatio-Temporal Data: An Analysis of urban Socioeconomic and Land Cover Change. *Transactions in GIS*, 9(1):5-17.

Mertens, B. and Lambin, E. F. 2000. Land-Cover Change Trajectories in Southern Cameroon. *Annuals of the Association of the American Geographers*, 90(3):467-497.

Mertens, K. C., Verbeke, L. P.C., Duchene, E. I., and De Wulf, R. R. 2003. Using Genetic Algorithms in Sub-Pixel Mapping. *International Journal of Remote Sensing*, 24(21): 4241-4247.

Messina, J. P., and Walsh, S. J. 2000. The Application of A Cellular Automata Model for Predicting Deforestation: Patterns and Processes of LULCC in the Ecuadorian, Amazon. 4th International Conference on Integrating GIS and Environmental Modeling (GIS /EM4): Problems, Prospects and Research Needs. September 2-8, 2000. Alberta, Canada.

Messina, J. P. 2001. *A Complex Systems Approach to Dynamic Spatial Simulation Modeling: LandUse and LandCover Change in the Ecuadorian Amazon (Dissertation).* Department of Geography, University of North Carolina, Chapel Hill, NC. 1-264.

Miller, A. B., Bryant, E. S., and Birnie, R. W. 1998. An Analysis of Land Cover Change in the Northern Forest of New England Using Multitemporal Landsat MSS Data. *International Journal of Remote Sensing*, 19(2): 245-265.

Mitchell, T. M., 1997. *Machine Learning*. McGraw-Hill, USA. 1-400.

Mortlock, D. F. 2004. Self-Organization and Cellular Automata Models. In: *Environmental Modeling: Finding Simplicity in Complexity*. Edited by: Wainwright, J. and Mulligan, M. John Wiley & Sons, Ltd. The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England. 349-367.

Mraz, M., Zimic, N., and Virant, J. 1999. Intelligent Bush Fire Spread Prediction Using Fuzzy Cellular Automata. *Journal of Intelligent and Fuzzy Systems*, 7(2):203-207.

MRLC (Multi-Resolution Land Characteristics Consortium). 1992. National Land Cover Dataset (NLCD 1992). http://www.mrlc.gov/, Last Accessed on May 1, 2005.

MRLC (Multi-Resolution Land Characteristics Consortium). 2001. National Land Cover Dataset (NLCD 2001). http://www.mrlc.gov/, Last Accessed on May 1, 2005.

Muller, M. R. and Middleton, J. 1994. A Markov Model of Land Use Change Dynamics in the Niagara region, Ontario, Canada, *Landscape Ecology*, 9(2):151-157.

Murata, T. 1989. Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE,* 177(4):541-580.

Nagendra, H., Munroe, D. K., and Southworth, J. 2004. From Pattern to Process: Landscape Fragmentation and the Analysis of Land Use/Land Cover Change. *Agriculture Ecosystems & Environment*. 101(2-3):111-115.

Neumann. J. V. 1966. *The Theory of Self-Reproducing Automata*. University of Illinois Press, Champaign, IL. 1-388.

NOAA (National Oceanic & Atmospheric Administration). 2006. *Regional Data: Coastal Louisiana Land Cover Change*. http://www.csc.noaa.gov/, Last Accessed on August 6, 2006.

Olague, G. 2002. Automated Photogrammetric Network Design Using Genetic Algorithms. *Photogrammetric Engineering & Remote Sensing*, 68(5):423-431.

Openshaw, S. 1988. Building an Automated Modeling System to Explore the Universe of Spatial Interaction Models. *Geographical Analysis*, 20(1):31-46.

Openshaw, S. 1992. Some Suggestions Concerning the Development of Artificial Intelligence Tools for Spatial Modeling and Analysis in GIS. *Annals of Regional Science*, 26(1):35-51.

Openshaw, S. 1995. Developing Automated and Smart Spatial Analysis Exploration Tools for GIS. *The Statistician*, 44(1):3-16.

Pal, M. and Mather, P. M., 2001. Decision Tree Based Classification of Remotely Sensed Data. The 22nd Asian Conference on Remote Sensing (AARS), November 5-9, 2001. Singapore.

Pal, S. K., Bandyopadhyay, S., and Murthy, C. A. 2001. Genetic Classifiers for Remotely Sensed Images: Comparison with Standard Methods. *International Journal of Remote Sensing*, 22(13):2545-2569.

Papagelis, A. and Kalles, D. 2000. GATree: Genetically Evolved Decision Trees. In: *Proceeding of ICTAI 2000.* 12th IEEE International Conference on Tools with Artificial Intelligence, November 13-15, 2000. Vancouver, British Columbia, Canada. 203-207.

Papagelis, A. and Kalles, D. 2001. Breeding Decision Trees Using Evolutionary Techniques. In: *Proceedings of ICML'01*. Edited by: Brodley, C. E., and Danyluk, A. P. 18th International Conference on Machine Learning. June 28-July 1, 2001. Williamstown, MA. Morgan Kaufmann, CA. 393-400.

Parker, D., Berger, T., and Manson, S. 2001. *Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop*. Indiana University. Indiana. 1-100.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., Deadman, P. 2003. Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. *Annals of the Association of American Geographers*, 93(2):314-337.

Parra, G. A., Mouchot, M. C., and Roux, C. 1996. A Multitemporal Land-Cover Change Analysis Tool Using Change Vector and Principal Components Analysis. In: *IGASS'96 Proceeding of International*. Geoscience and Remote Sensing Symposium, May 27-31, 1996. Lincoln, NE. 3:1753-1755.

Patricia, I. R., Victor, S. C., Peterson, A. T.. 2004. Predicting Distributions of Mexican Mammals Using Ecological Niche Modeling. *Journal of Mammalogy*, 85(4):658-662.

Peddle, D. R., Johnson, R. L., Cihlar, J., Guindon, B., and Latifovic, R. 2004. Large Area Forest Classification and Biophysical Parameter Estimation Using the 5-Scale Canopy Reflectance Model in Multiple-Forward-Model. *Remote Sensing of Environment*, 89(2):252-263.

Peimann, C. J. 1988. Modeling Hospital Information Systems with Petri Nets. *Methods of Information in Medicine*. 27(1):17-22.

Penland, S., Beall, A., and Kindinger, J. 2002. Environmental Atlas of the Lake Pontchartrain Basin. In: *U.S. Geological Survey Open File Report 02-xxx, CD-ROM*. Lake Pontchartrain Basin Foundation, New Orleans, LA. http://coastal.er.usgs.gov/, Last accessed on May 6, 2007.

Pereira, A. G. 2001. Evolutionary Modeling of Route: The Case of Road Design. In: *Spatial Evolutionary Modeling*. Edited by: Krzanowski, R. and Raper, J. Oxford University Press, Inc., New York, NY. 180-202.

Perkins, S. J., Theiler, J., Brumdy, S. P., Harvey, N. R., Porter, R., Szymanski, J. J., and Bloch, J. J. 2000. GENIE: A Hybrid Genetic Algorithm for Feature Classification in Multispectral Images. *Proc. SPIE*. 4120:52-62.

Peterson, J. L. 1981. *Petri Net Theory and the Modeling of Systems*. Prentice-Hall, Inc, NJ. 1-78.

Petit, C., Scudder, T., and Lambin, E. 2001. Quantifying Processes of Land-Cover Change by Remote Sensing: Resettlement and rapid Land-Cover Changes in South-Eastern Zambia. *International Journal of Remote Sensing*, 22(17): 3435-3456.

Phillips, O. L., Baker, T. R., Arroyo, L., Higuchi, N., Killeen, T. J., Laurance, W. F., Lewis, S. L., Lioyd, J., Malhi, Y., Monteagudo, A., Neill, D. A., Vargas, P. N., Silva, J. N. M., Terborgh, J., Martinez, R. V., Alexiades, M., Almeida, S., Brown, S., Chave, J., Comiskey, J. A., Czimczik, C. I., Fiore, A. D., Erwin, T., Palacios, W., Patino, S., Pitman, N. C. A., Quesada, C. A., Saldias, M., Lezama, A. T., and Vinceti, B. 2004. Pattern and Process in Amazon Tree Turnover, 1976-2001. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 359(1443):381-407.

Pijanowski, B. C., Long, D. T., Gage, S. H., and Cooper, W. E. 1997. *A Land Transformation Model: Conceptual Elements Spatial Object Class Hierarchy, GIS Command Syntax and an Application for Michigan's Saginaw Bay Watershed*. http://www.ncgia.uscb.edu/. Last accessed on May 12, 2006.

Pijanowski, B. C., Shellito, B., Pithadia, S., and Alexdandridris, K. 2002. Forecasting and Assessing the Impact of Urban Sprawl in Coastal Watersheds Along Eastern Lake Michigan. *Lake & Reservoirs: Research & Management.* 7(3):271-285.

Pinter, S.S. and Golani, M. 2004. Discovering Workflow Models from Activities' Lifespans. *Computers in Industry*, 53(3):283-296.

Prakash, A., and Gupta, R. P. 1998. Land-Use Mapping and Change Detection in Coal Mining Area – A Case Study in the Jharia Coalfield, India. *International Journal of Remote Sensing*, 19(3): 391-410.

Price, K. O., Pyke, D. A., and Mendes, L. 1992. Shrub Dieback in a Semiarid Ecosystem: the Integration of Remote Sensing and GIS for Detecting Vegetation Change. *Photogrammetric Engineering and Remote Sensing*, 58(4):455-463.

Putman, S. H. 1995. EMPAL and DRAM Location and Land Use Models: A technical Overview. In: *Proceedings of the Land Use Modeling Conference*, February 19-21, 1995. Dallas, TX.

Read, J. M., and Lam, N. S-N. 2002. Spatial Methods for Characterizing Land Cover and Detecting Land Cover Changes for the Tropics. *International Journal of Remote Sensing*, 23(12): 2457-2474.

Reddy, J. P., Kumanan, S. and Chetty, O. V. K. 2001. Application of Petri Nets and a Genetic Algorithm to Multi-Mode Multi-Resource Constrained Project Scheduling. *The International Journal of Advanced Manufacturing Technology*, 17(4):305-314.

Richards, J. A. and Jia X. P. 1993. *Remote Sensing Digital Image Analysis: An Introduction* (2nd). Springer-Verlag, Berlin, Germany. 1-429.

Ridd, M. K., and Liu, J. 1998. A Comparison of Four Algorithms for Change Detection in an Urban Environment: A Remote Sensing Perspective. *Remote Sensing of Environment*, 63(2):95-100.

Rocha, M. and Neves, J. 1999. Preventing Premature Convergence to Local Optima in Genetic Algorithms via Random Offspring Generation. In: *Lecture Notes in Computer Science*. Springer-Verlag, New York, NY. 1611:127-136.

Rodrigo, R. D., and Nicholls, M. G. 1998. Modeling real World Industrial Systems : an Inter-Disciplinary Approach based on Petri Nets. *Mathematics and Computers in Simulation*, 44(6):587-597.

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L. 2003. Land-Cover Change Monitoring with Classification Trees Using Landsat TM and Ancillary Data. *Photogrammetric Engineering &Remote Sensing*, 69(7):793-804.

Rounsevell, M. D. A., Evans, S. P., Mayr, T. R., and Audsley, E. 1997. Integrating Biophysical and Socio-Economic Models for Land Use Studies. In: *Proceedings of the ITC & ISSS Conference on Geo-Information for Sustainable Land Management*, August 17-21, 1997. Enschede, Netherlands.

Rudolph, G. 1994. Convergence Analysis of Canonical Genetic Algorithms. *IEEE Transactions on Neural Networks*, 5(1):96-101.

Samadzadegan, F., Azizi, A., and Abootalebi, A. 2005. Automatic Determination of the Optimum Generic Sensor Model Based on Genetic Algorithm Concepts. *Photogrammetric Engineering and Remote Sensing*, 71(3):277-288.

Schimm, G. 2004. Miming Exact Models of Concurrent Workflows. *Computers in Industry*, 53(3):265-281.

Schneeweiss, W. G. 2004. *Petri Net Picture Book*. LiLoLe-Verlag GambH Publ. Co. Ltd. Hagen, Germany. 1-113.

Schowengerdt, R. A. 1996. On the Estimation of Spatial-Spectral Mixing with Classifier Likelihood Functions. *Pattern Recognition Letters*, 17(13):1379-1387.

Schwardy, E. 2003. *Optimization of Petri Nets Structure Using Genetic Programming*. In: *Proceeding of SAMI 2002*. 1st Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence (SAMI 2003), February 12-14, 2003. Herlany, Slovakia.

Serpico, S. B., and Bruzzone, L. 1999. Change Detection. In: *Information Processing for Remote Sensing*. Edited by: Chen, C. H. World Scientific Publishing, Singapore. 319-336.

Seto, K. C., Woodcock, C. E., Song, C., Huang, X., Lu, J., and Kaufmann, R. K. 2002. Monitoring Land-Use Change in the Pearl River Delta Using Landsat TM. *International Journal of Remote Sensing*, 23(10):1985-2004.

Settle, J. and Campbell, N. 1998. Linear Mixing and the Estimators of Subpixel Fractional Cover When Mixing is Linear. *IEEE Transactions and Geosciences and Remote Sensing*. 36:163-169.

Shaban, M. A., and Dikshit, O. 2001. Improvement of Classification in Urban Areas by the Use of Textural Features: the Case Study of Lucknow City, Uttar Pradesh. *International Journal of Remote Sensing*, 22(4):565-593.

Shalan, M. A., Arora, M. K., and Ghosh, S. K. 2003. An Evaluation of Fuzzy Classification from IRS 1C LISS III Imagery: a Case Study. *International Journal of Remote Sensing*, 24(15):3179-3186.

Simard, M., DeGrandi, G., Saatchi, S., nd Mayaux, P. 2002. Mapping Tropical Coastal Vegetation Using JERS-1 and ERS-1 Radar Data with a Decision Tree Classifier. *International Journal of Remote Sensing*, 23(7):1461-1474.

Singh, A. 1986. Change Detection in the Tropical Forest Environmental of Northeastern India Using Landsat Remote Sensing and Tropic Land Management. In: *Remote Sensing and Tropical Land Management*. Edited by Eden, M. J., and Parry, J. T. John Wiley & Sons, New York, NY. 237-254.

Singh, A. 1989. Digital Change Detection Techniques Using Remotely Sensed Data. *International Journal of Remote Sensing*, 10(10):989-1003.

Singh, A. K. 2003. *Modeling Land Use Land Cover Changes Using Cellular Automata in a Geo-Spatial Environment (Thesis).* International Institute for Geo-Information Science and Earth Observation, Enschede, Netherlands. 1-47.

Sirakoulis, G. C., Karafyllidis, I., and Thanailakis, A. 2000, A Cellular Automata Model for the Effects of Population Movement and Vaccination and Epidemic Propagation. *Ecological Modeling*, 133(3):209-223.

Slater, J., and Brown, R. 2000. Changing Landscapes: Monitoring Environmentally Sensitive Areas Using Satellite Imagery. *International Journal of Remote Sensing*, 21(13-14):2753-2767.

Sloot, Peter, Chen, F., and Boucher, C. 2002. Cellular Automata Model of Drug Therpy for HIV Infection. In: *Lecture Notes in Computer Science*. Springer-Verlag, New York, NY. 2493:282-293.

Soares, V. P., and Hoffer, R. M. 1994. Eucalyptus Forest Change Classification Using Multi-Date Landsat TM Data. *Proc. SPIE*, 2314:281-291.

Sohl, T. 1999. Change Analysis in the United Arab Emirates: an Investigation of Techniques. *Photogrammetric Engineering and Remote Sensing*, 65(4):475-484.

Sondgerath, D., and Schroder, B. 2004. Population Dynamics and Habitat Connectivity Affecting the Spatial Spread of Populations – a Simulation Study. *Landscape Ecology*. 17(1):57-70.

Song, X. M., Fan, G. L., and Rao, M. 2005. Automatic CRP Mapping Using Nonparametric Machine Learning Approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):888-897.

Srinivas, M and Patnaik, L. M. 1994. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 24(4):656-666.

Srinivasan, A., and Richards, J. A. 1990. Knowledge-based Techniques for Multisource Classification. *International Journal of Remote Sensing*, 11(3):505-525.

Stathakis, D. and Vasilakos, A. 2006. Comparison of Computational Intelligence based Clssification Techniques for Remotely Sensed Optical Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2305-2318.

Stewart, T. J., Janssen, R., and van Herwijnen, M. 2004. A Genetic Algorithm Approach to Multiobjective Land Use Planning. *Computers & Operations Research*, 31(14):2293-2313.

Stockman, A. K., Beamer, D. A. and Bond, J. E. 2006. An Evaluation of a GARP Model as an Approach to Predicting the Spatial Distribution of Non-Vagile Invertebrate Species. *Diversity and Distribution*, 12(1):81-89.

Stockwell, D., and Peters, D. 1999. The GARP Modeling System: Problems and Solutions to Automated Spatial Prediction. *International Journal of Geographic Information Science*, 13(2):143-158.

Stockwell, D. R. B., Beach, J. H., Stewart, A., Voronstov, G., Vieglais, D., Pereira, R. S. 2006. The Use of the GARP Genetic Algorithm and Internet Grid Computing in the Lifemapper World Atlas of Species Biodiversity. *Ecological Modeling*, 195(1-2):139-145.

Stuckens, J., Coppin, P. R., and Bauer, M. E. 2000. Integrating Contextual Information with Per-pixel Classification for Improved Land Cover Classification. *Remote Sensing of Environment*, 71(3):282-296.

Sui, D. Z. and Zeng, H. 2001. Modeling the Dynamics of Landscape Structure in Asia's Emerging Desakota Regions: A Case Study in Shenzhen. *Landscape and Urban Planning*, 53(1-4):37-52.

Sunar, F. 1998. An Analysis of Change in a Multi-Date Data Set: A Case Study in the Ikitelli Area, Istanbul, Turkey. *International Journal of Remote Sensing*, 19(2):225-235.

Suri, R. 1985. An Overview of Evaluative Models for Flexible Manufacturing Systems. *Annals of Operations Research*, 3(1):13-21.

Syphard, A. D., Clarhe, K. C., and Franklin, J. 2005. Using a Cellular Automata Model to Forecast the Effects of Urban Growth on Habitat Pattern in Southern California. *Ecological Complexity*, 2(2):185-203.

Taylor, J. C., Brewer, T. R., and Bird, A. C. 2000. Monitoring Landscape Change in the National Park of England and Walse Using Aerial Photo Interpretation and GIS. *International Journal of Remote Sensing*, 21(13-14):2737-2752.

Tobler, W. R. 1979. Cellular Geography. In: *Philosophy in Geography*. Edited by: Gale, S. and Olsson, G. Reidel Publishing Company, Dordrecht, Netherlands. 379-386.

Toll, D. 1985. Effect of Landsat Thematic Mapper Sensor Parameters on Land Cover Classification. *Remote Sensing of Environment*, 17(2):129-140.

Torrens, P. M. 2000. *How Cellular Models of Urban Systems Work*. http://www.casa.ucucl.ac.uk/publications/. Last accessed on March 4, 2007.

Torrens, P. M. 2001. *Can Geocomputation Save urban Simulation? Throw Some Agent into the Maxture, Simmer, and Wait*. http://www.casa.ucucl.ac.uk/publications/. Last accessed on March 4, 2007.

Tottrup, C. 2004. Improving Tropical Forest Mapping Using Multi-Date Landsat TM Data and Pre-Classification Image Smoothing. *International Journal of Remote Sensing*, 25(4):717-730.

Tso, B. C. K. and Mather, P. M. 1999. Classification of Multisource Remote Sensing Imagery Using Genetic Algorithm and Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1255-1260.

Tso, B. C. K. and Mather, P. M. 2001. *Classification Methods for Remotely Sensed Data*. Taylor & Francis, New York, NY. 1-326.

Turner, M. G., Costanza, R. and Sklaar, F. H. 1989. Methods to Evaluate the Performance of Spatial Simulation Models. *Ecological Modeling*, 48(1-2):1-18.

Turner, M. G. 1987. Spatial Simulation of Landscape Changes in Georgia: A Comparison of Transition Models. *Landscape Ecology*, 1(1):29-36.

Turner, M. G., 1989. Landscape Ecology: The Effect of Pattern on Process. *Annual Review of Ecology and Systematics*, 20:171-197.

U.S. Census Bureau. 2006. *State & County Quick Facts*, http://quickfacts.census.gov. Last Accessed on May 6, 2007.

U.S. EPA. 2000. *Projecting Land-Use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns. EPA/600/R-00/098*. U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, OH. 1-260.

U.S. EPA. 2001. *Better Assessment Science Integrating Point & Nonpoint Sources (BASINS) Version 3.0*. http://www.epa.gov/waterscience/basins/, Last Accessed on May 1, 2005.

U.S. EPA. 2002. *Assessment Data for Louisiana, Tickfaw Watershed, Year 2002*. http://oaspub.epa.gov. Last Accessed on May 6, 2007.

U.S. EPA. 2005. *Surf Your Watershed*. http://cfpub.epa.gov/. Last Accessed on May 5, 2006.

U.S. EPA. 2007. *Smart Growth Index 2.0*. http://www.epa.gov/smartgrowth/. Last Accessed on March 4, 2007.

USGS. 2005. *Free Landsat Image from USGS*. http://geology.com/news/2005/, Last accessed on May 1, 2006.

van der Aalst, W. M. P. 1994. Putting Petri Nets to Work in Industry, *Computer in Industry*, 25(1):45-54.

van der Aalst, W. M. P., and van Hee, K. M. 1996. Business Process Redesign: A Petri-net-based Approach. *Computers in Industry*, 29(1-2):15-26.

van der Aalst, W. M. P. 1997. Application and Theory of Petri Net in Verification of Workflow Nets. *Lecture Notes in Computer Science*, 1248:407-426.

van der Aalst, W. M. P. 1998. The Application of Petri Nets to Workflow Management. The *Journal of Circuits, Systems and Computers*, 8(1):21-66.

van der Aalst, W. M. P. 2002. Making Work Flow: On the Application on Petri nets to Business Process Management. In: *Proceedings of Applications and Theory of Petri Nets 2002: 23rd International Conference, ICATPN 2002*, June 24-30, 2002. Adelaide, Australia.

van der Aalst, W. M. P. 2003. Business Process Management: A Survey, *BPM 2003, LNCS*, 2678:1-12, Springer-Verlag, New York, NY.

van der Aalst, W. M. P., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., Weijters, A. J. M. M. 2003. Workflow Mining: a Survey of Issues and Approaches. *Data & Knowledge Engineering*, 47(2):237-267.

van der Aalst, W., Weijters, T., and Maruster, L. 2004. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128-1141.

van der Aalst, W. M. P., and Ton Weijters, A. J. M. M. 2005. Process Mining. In: *Process-Aware Information System*. Edited by: Dumas, M., van der Aalst, W., and Ter Hofstede, A. H. M. John Wiley & Sons, Inc., Hoboken, NJ. 235-253.

van der Aalst, W. M. P., Alves de Medeiros, A. K., and Weijters, A. J. M. M. 2005. Genetic Process Mining. In: Lecture Notes in Computer Science, Edited by: Ciardo, G. and Darondeau, P. Springer-Verlag, New York, NY. 3536:48-69.

Veldkamp, A. and Fresco, L. O. 1996a. CLUE: A Conceptual Model to Study the Conversion of Land Use and Its Effects. *Ecological Modeling*, 85(2):253-270.

Veldkamp, A. and Fresco, L. O. 1996b. CLUE-CR: An Integrated Multi-Scale Model to Simulate Land Use Change Scenarios in Costa Rica. *Ecological Modeling*. 91(1):231-248.

Veldkamp, A., and Lambin, E. F. 2001. Predicting Land-Use Change. *Agriculture Ecosystems & Environment*. 85(1):1-6.

Verbeke, L. P. C., Vabcoillie, F. M. B., and De Wulf, R. R. 2004. Reusing Back-Propagating Artificial Neural Network for Land Cover Classification in Tropical Savannahs. *International Journal of Remote Sensing*, 25(14):2747-2771.

Voinov, A., Voinov, H., and Costanza, R. 1999a. Surface Water Flow in Landscape Models: Patuxent Watershed Case Study. *Ecological Modeling*. 119(2-3):211-230.

Voinov, A., Costanza, R., Wainger, L., Boumans, R., Maxwell, T., and Voinov, H. 1999b. Patuxent Landscape Model: Integrated Ecological Economic Modeling of a Watershed. *Environmental Modeling and Software*, 14(5):473-491.

Waddell, P. 2002. UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of American Planning Association*, 68(3):297-314.

Waddell, P. and Ulfarsson, G. F. 2004. Introduction to Urban Simulation: Design and Development of Operations. In: *Handbook of Transport, Volume 5: Transport Geography and Spatial Systems*. Edited by: Hensher, D. A., Button, K. J., Haynes, K. E., Stopher, P., Elsevier Science, San Diego, CA. 203-236.

Wall, M. 1996. *A C++ Library of Genetic Algorithm Components*, http://lancet.mit.edu/ga/. Last access at Dec. 31, 2006.

Wang, S. and Nakayama, K. 2005. An Event-based Spatiotemporal Approach. *Transactions on Computer and Information Theory*, 1(1):15-23.

Wang, X. P. and Cao, L. M. 2002. *Genetic Algorithms: Theory, Application, and Software Implementation*. Shiang JiaoTong University Press, Shiang, China. 1-339.

Wegener, M. 1998. *The IRPUD Model: Overview*, http://www.raumplanung.uni-dortmund.de/. Last accessed on May 6, 2007.

Wegener, M. 2000. Spatial Models and GIS. In: *Spatial Models with GIS: New and Potential Models*, Edited by: Fortheringham, S. and Wegener, M. Taylor and Francis, Landon. 3-20.

Weijters, T. and Paredis, J. 2002. Genetic Rule Induction at an Intermediate Level. *Knowledge-Based System*, 15(1):85-94.

White, R. and Engelen, G. 1993. Cellular Automata and Fractal Urban Form: A Cellular Modeling Approach to the Evolution of Urban Land-Use Patterns. *Environment and Planning A*, 25(8):1175-1199.

White, R. and Engelen, G. 1994. Cellular Dynamics and GIS: Modeling Spatial Complexity. *Geographical Systems*, 1(2):237-253.

White, R. and Engelen, G. 1997. Cellular Automata as the Basis of Integrated Dynamic Regional Modeling. *Environment and Planning B*, 24 (2):235-246.

White, R. and Engelen, G. 2000. High-Resolution Integrated Modeling of the Spatial Dynamics of Urban and Regional Systems. *Computers, Environment and Urban Systems*, 24(5):383-400.

Will, J. and Heiner, M. 2002. *Petri Nets in Biology, Chemistry, and Medicine – Bibliography*, http://www-dssz.informatik.tu-cottbus.de/, Last accessed on March 4, 2007.

Williams, D. L., Irins, J. R., Markham, B. L., Nelson, R. F., Toll, D. L., Latty, R. S., and Stauffer, M. L. 1984. A Statistical Evaluation of the Advantages of Landsat Thematic Mapper Data in Comparison to Multispectral Scanner Data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(3):294-302.

Willox, R., Grammaticos, B., Carstea, A. S., and Ramani, A. 2003. Epidemic Dynamics: Discrete-Time and Cellular Automata Models. *Physica A: Statistical Mechanics and Its Applications*, 328(1-2):13-22.

Witten, I. H. and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, CA. 1-560.

Wolfram, S. 1984. Cellular Automata: A Model of Complexity, *Nature*, 311:419-424.

Wolfram, S. 2002. *A New Kind of Science*, Wolfram Media, Inc. Champaign, IL. 1-1197.

Wolter, P. T., Host, G. E., Mladenoff, D. J., and Crow, T. R. 1995. Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery. *Photogrammetric Engineering and Remote sensing*. 61(9):1129-1143.

Wu, F. 1998a. Simland: A Prototype to Simulate Land Conversion through the Integrated GIS and CA with AHP-Derived Transition Rules. *International Journal of Geographic Information Science*, 12(1):63-82.

Wu, F. 1998b. Simulating Urban Encroachment on Rural Land with Fuzzy-Logic-Controlled Cellular Automata in a Geographical Information System. *Journal of Environmental Management*, 53(4):293-308.

Wu, F. 1999. GIS-based Simulation as an Exploratory Analysis for Space-Time Processes. *Journal of Geographical System*, 1(3):199-218.

Wu, F. 2002. Calibration of Stochastic Cellular Automata: the Application to Rural-urban land Conversions. *International Journal of Geographical Information Science*. 16(8):759-818.

Wu, F. and Webster, C. J. 1998. Simulation of Land development through the Integration of Cellular Automata and Multicriteria Evaluation. *Environment and Planning B*, 25(1):103-126.

Wu, F., and Martin, D., 2002, Urban Expansion Simulation of Southeast England Using Population Surface Modeling and Cellular Automata. *Environment and Planning A,* 34(10): 1855-1876.

Wu, K. S. 2005. *Long-Term Freshwater Input and Sediment Load from Three Tributaries to Lake Pontchartrain (Dissertation).* School of Renewable Nature Resource, Louisiana State University. Baton Rouge, LA, 1-138.

Wu, Q. H. and Cao, Y. J. 1997. Stochastic Optimization of Control parameters in Genetic Algorithms. *IEEE International Conference on Evolutionary Computation*, April 13-16, 1997. Indianapolis, IN. 77-80.

Wu, S., Lai, C. and Sun, Y. M. 2006. Applying Data Mining and Petri Net in Reengineering of Manufacture Management Information System, *2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06)*. 200-206.

Wu, Z. H. 2006. *Introduction to Petri Net (in Chinese).* China Machine Press, Beijing, China. 1-300.

Xiao, N. C., Bennett, D. A. and Armstrong, M. P. 2002. Using Evolutionary Algorithms to Generate Alternative for Multiobjective Site-Search Problems. *Environment and Planning A*, 34(4):639-656.

Xie, Y. 1996. A Generalized Model for Cellular Urban Dynamics. *Geographical Analysis*, 28(4):350-373.

Yang, Q. W., Jiang, J. P., and Chen, G. 2000. How to Select Optimal Control Parameters for Genetic Algorithms. In: ISIE' 2000 *Proceeding,* 1:37-41. IEEE International Symposium on Industrial Electronics, December 4-8, 2000. Cholula, Puebla, Mexico.

Yang, X. and Lo, C. P. 2002. Using a Time Series of Satellite Imagery to Detect Land Use and Land Cover Change in the Atlanta, Georgia Metropolitan Area. *International Journal of Remote Sensing*, 23(9):1775-1798.

Yeh, A. G. O., and Li, X., 2000. Simulation of Compact Cities base on the Integration of Cellular Automata and GIS. In: *Theoretical and Practical Issues on Cellular Automata*. Edited by: Bandini, S., and Worsch, T. Springer-Verlag, London. 170-178.

Yeh, A. G. O., and Li, X. 2001. A Constructrained CA Model for the Simulation and Planning of Sustainable Urban Forms by Using GIS. *Environment and Planning B*, 28 (5):733-753.

Yeh, A. G. O., and Li, X., 2002. A Cellular Automata Model to Simulate Development Density for Urban Planning. *Environment and Planning B*, 29(3):431-450.

Yeh, G. O., and Li, X. 2003. Simulation of Development Alternative Using Neural Networks, Cellular Automata, and GIS for urban Planning. *Photogrammetric Engineering and Remote Sensing*, 69(9):1043-1052.

Yen, H. 2002. Sequential Versus Concurrent Languages of Labeled Conflict-Free Petri Net. *IEEE Transaction on Automatic Control*, 47(7):1158-1162.

Yi, X. C. and Kochut, K. J. 2005. JCPNet Tool and Automated Analysis of Distributed System, ACM Southeast Regional Conference, Kennesaw, Georgia. In: *Proceedings of the 43rd Annual Southeast Regional Conference*. ACM Press, New York, NY. 2:327-332.

Yin, Z. C., and Li, L. 2004. Petri Nets-based Spatio-Temporal Data Model. *Geomatics and*

Information Science of Wuhan University, 29(8):740-743.

Yodis, G., Colten, C. and Johnson, D. C. 2003. Geography of Louisiana. McGraw-Hill. USA. 1-207.

Zhang, Q., Wang, J., Peng, X., Gong, P., and Shi, P. 2002. Urban Build-up Land Change Detection with Road Density and Spectral Information from Multitemporal Landsat TM Data. *International Journal of Remote Sensing*, 23(15):3057-3078.

Zhao, F., and Chung, S. 2006. *A Study of Alternative Land Use Forecasting Models*. FDOT Systems Planning Office, Tallahassee, FL. 1-180.

Zhou, J., and Civco, D. L. 1996. Using Genetic Learning Neural Networks for Spatial Decision Making in GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11):1287-1295.

Zhou, J., and Civco, D. L. 1997. Using Genetic Learning Neural Networks for Feature Extraction and Classification. In: *Proc. ASPRS/ACSM Annual Convention*, Seattle, WA. 95-105.

VITA

Fei Wang was born in Wuhan, China, in 1961. He received his bachelor degree in biology from Wuhan Teacher's College in 1984. After then, he had worked in Hubei Environmental Protection Institute for ten years. He received his first master's degree in environmental geography from Beijing Normal University in 1995, and a second master's degree in computer science from Southern University, Baton Rouge, Louisiana, in 2002. Since June 2003, he has pursued a doctoral degree in geography, focusing on geographic information systems and remote sensing, at Louisiana State University, Baton Rouge. He is expected to receive the degree of Doctor of Philosophy in Fall 2007.