Integrating Artificial Neural Networks, Image Analysis and GIS for Urban Spatial Growth Characterization

Libin Zhou
INTEGRATING ARTIFICIAL NEURAL NETWORKS, IMAGE ANALYSIS AND GIS FOR URBAN SPATIAL GROWTH CHARACTERIZATION

By
LIBIN ZHOU

A dissertation submitted to the Department of Geography in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Degree Awarded:
Spring Semester, 2012
Libin Zhou defended this dissertation on March 14, 2012.
The members of the supervisory committee were:

Xiaojun Yang
Professor Directing Dissertation

Timothy S Chapin
University Representative

James B Elsner
Committee Member

Jon A Stallins
Committee Member

Trajco V Mesev
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.
To my parents, Shiyu and Yuezhen, and my husband and daughter, Qiguang and Jialu
ACKNOWLEDGEMENT

I am grateful to my committee for their advice and supports throughout my study at the Florida State University (FSU). They are professors Xiaojun Yang, James B. Elsner, Jon A. Stallines, Trajco V. Mesev, and Timothy S. Chapin. I particularly thank Dr. Yang. Without his encouragement, I would have not completed this dissertation. I also want to thank the Department of Geography at the Florida State University for offering me four-year assistantship. In addition, I want to acknowledge the China’s State Key Laboratory of Urban and Regional Ecology, and the Chinese Academy of Sciences for their financial supports on my field trips in Beijing, China. Thanks also go to Dr. Liding Chen for his help in many ways.
# TABLE OF CONTENTS

ACKNOWLEDGEMENT ................................................................................................................................. iv

LIST OF TABLES ............................................................................................................................................... vii

LIST OF FIGURES ........................................................................................................................................ viii

ABSTRACT ..................................................................................................................................................... x

CHAPTER ONE: INTRODUCTION ................................................................................................................... 1

  1.1 Research background .......................................................................................................................... 1

  1.2 Research objectives ........................................................................................................................... 4

  1.3 Selection of the study area ................................................................................................................ 5

  1.4 Research approach and hypotheses ................................................................................................. 6

  1.5 Organization of the dissertation ....................................................................................................... 9

CHAPTER TWO: LITERATURE REVIEW ........................................................................................................ 11

  2.1 Introduction ......................................................................................................................................... 11

  2.2 Complexity theory ............................................................................................................................ 11

  2.3 Remote sensing ............................................................................................................................... 12

  2.4 Artificial neural networks ................................................................................................................ 13

  2.5 Geographic information systems .................................................................................................... 15

CHAPTER THREE: RESEARCH METHODOLOGY ....................................................................................... 19

  3.1 Introduction ....................................................................................................................................... 19

  3.2 Study site .......................................................................................................................................... 22

  3.3 Data acquisition and collection ....................................................................................................... 29

  3.4 Urban land change mapping .......................................................................................................... 33

  3.5 Urban growth characterization ....................................................................................................... 35

CHAPTER FOUR: ARTIFICIAL NEURAL NETWORKS FOR LAND CLASSIFICATION ....................... 37

  4.1 Introduction ....................................................................................................................................... 37

  4.2 Fundamentals of ANNs .................................................................................................................... 37

  4.3 Internal parameters and classification accuracy .............................................................................. 57

  4.4 Training algorithm performance .................................................................................................... 66
4.5 Neural network architectures assessment ................................................................. 71
4.6 Conclusions ................................................................................................................. 77

CHAPTER FIVE: URBAN GROWTH CHARACTERIZATION ........................................... 83
5.1 Introduction .................................................................................................................. 83
5.2 Urban land change mapping ...................................................................................... 83
5.3 Urban growth characterization .................................................................................. 96
5.4 Spatial and temporal dynamics of urban growth ....................................................... 111
5.5 Conclusions ................................................................................................................ 125

CHAPTER SIX: SUMMARY AND CONCLUSIONS ......................................................... 128
REFERENCES ................................................................................................................... 132
BIOGRAPHICAL SKETCH .............................................................................................. 143
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>List of satellite images used for land use/land cover mapping</td>
<td>31</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>List of some neural networks in machine learning</td>
<td>39</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>List of some training algorithms that are particularly tied with a feed-forward network</td>
<td>51</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>List of the internal network parameters tested</td>
<td>58</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>List of the 53 neural network models configured with various internal parameter settings</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Comparison of classification accuracies by artificial neural networks (ANNs, Model 48 of Table 4.4) and Gaussian Maximum Likelihood (GML) classifier</td>
<td>65</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Summary of the training algorithm performance</td>
<td>69</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Experimental results of land cover classification by SOM neural networks as related to the map size and fine tuning iterations</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>Experimental results of land cover classification by Fuzzy ARTMAP neural networks as related to various settings for the three parameters: choice, vigilance, and learning rate</td>
<td>75</td>
</tr>
<tr>
<td>Table 4.9</td>
<td>Experimental results of land cover classification by the probabilistic neural networks as related to varying spread constants</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.10</td>
<td>Comparisons of the optimal performances of the four types of neural networks for land cover classification</td>
<td>78</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Land cover classification scheme</td>
<td>86</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Confusion matrix and Kappa statistics for the 2000 land cover map</td>
<td>88</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Confusion matrix and Kappa statistics for the 2009 land use/cover map</td>
<td>89</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Statistics of the urban growth in the Beijing metropolitan area during 2000-2009 at metropolitan and functional zone levels, respectively</td>
<td>94</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Six selected metrics for landscape pattern analysis</td>
<td>99</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>Results of the landscape metrics analysis with cell sizes of 30m, 50m, and 100m, respectively, based on the 2009 urban extent map</td>
<td>101</td>
</tr>
<tr>
<td>Table 5.7</td>
<td>Results of the landscape metrics analysis at the metropolitan level in 2000 and 2009</td>
<td>106</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1  Theoretical and technological integration of the pattern-process analysis in support of urban growth research................................................................. 8
Figure 3.1  The major working phases in the dissertation research..................................... 20
Figure 3.2  Working procedural route................................................................................ 21
Figure 3.3  Location of the case study site........................................................................... 23
Figure 3.4  The population growth in the Beijing metropolitan area.................................... 25
Figure 3.5  Trends of gross domestic product (GDP) in the Beijing metropolitan area........... 27
Figure 3.6  Industrial structure evolution in the Beijing metropolitan area: 1978-2008. ......... 28
Figure 3.7  Routes of ground-truth survey in the Beijing metropolitan area......................... 32
Figure 3.8  Landsat TM images and relevant photos showing different land use/land cover types................................................................................................................. 34
Figure 4.1  A fully-connected multilayer perceptron (MLP) neutral network with a 4 x 5 x 4 x 2 structure.................................................................................................................. 40
Figure 4.2  A Kohonen’s SOM neutral network with a two-dimension output layer.............. 42
Figure 4.3  Structure of ARTMAP neural networks............................................................ 43
Figure 4.4  Structure of probabilistic neural networks (PNNs)............................................. 45
Figure 4.5  Curves of activation functions as related to different threshold values............. 49
Figure 4.6  Classification accuracies as related to different internal settings........................ 63
Figure 4.7  Illustration of the optimal performances of four types of neural networks in urban land cover classification. .......................................................................................... 79
Figure 4.8  A systematic approach to image classification by neural networks..................... 80
Figure 5.1  Working flow chart for image preprocessing........................................................ 85
Figure 5.2  Visualizing the classification accuracies by comparing the image data with the classified land cover map............................................................................................................. 90
Figure 5.3  The 2000 and 2009 urban extent maps............................................................... 92
Figure 5.4  Urban spatial growth during 2000-2009 in the Beijing metropolitan area............. 93
Figure 5.5  Trends of the urban growth at various functional zones from 2000 to 2009. ....... 95
Figure 5.6  Trends of landscape metrics in 2009, namely, NP, PMA, LSI, MPAR, M-ENN and AI, as related to various cell sizes ........................................................................................................... 102
Figure 5.7  Results of the landscape metrics analysis at the functional zone level in 2000... 104
Figure 5.8  Results of the landscape metrics analysis at the functional zone level in 2009... 105
Figure 5.9  Trends of urban growth patterns at different functional zones during 2000-2009, indicated by landscape metrics analysis................................................................. 109
Figure 5.10 The analytic procedures for urban growth modeling................................. 112
Figure 5.11 Illustration of the landscape metrics analysis with moving windows............. 113
Figure 5.12 Results of the landscape metrics analysis with moving windows for 2000.... 115
Figure 5.13 Results of the landscape metrics analysis with moving windows for 2009.... 116
Figure 5.14 Results of the contour analysis: NP, PMA, LSI and AI................................... 118
Figure 5.15 The three universal zones identified through an overlay analysis for 2000...... 120
Figure 5.16 The three universal zones identified through an overlay analysis for 2009...... 121
Figure 5.17 Comparisons of the recognized universal zones between 2000 and 2009..... 124
ABSTRACT

Outward urban growth, driven by increasing population, economic development, and technological advancement, has become a worldwide phenomenon. Such growth is often viewed as the vitality of a regional economy. But it has brought negative impacts on the environment such as biodiversity loss, soil erosion, hydrological perturbation, water and solid pollution, and global warming. Monitoring and modeling urban spatial growth are important for environmental sustainability and urban planning.

This dissertation research has aimed at the investigation of urban growth patterns, urban growth processes, and their relevance through the lens of complexity theory to improve our understanding of the spatial and temporal dynamics of urban growth in a rapidly growing metropolitan area. Central to this research effort is the development of a technological framework that tightly integrates satellite imagery processing, artificial intelligence, and geographic information systems (GIS). Specifically, this project includes two principle components. One is to examine the use of artificial neural networks for improving urban land cover change detection from remote sensor data. Due to their capability of dealing with nonlinear and complex phenomena, integrating artificial neural networks with remote sensing has improved the performance of image classification for the fragmented and heterogeneous landscape in an urban environment. The other component is to characterize urban spatial growth at the metropolitan, functional zone, and cell levels by using three approaches: urban land change mapping, landscape metrics analysis, and moving windows analysis. This part of the research has provided insights into urban growth dynamics in urban societies that are not comparable to either industrial or post-industrial cities in the United States through measuring the spatial and temporal variations of urban patterns and processes at different scales. These societies have unique urban forms and development trajectories due to technological robustness and contemporary international and domestic socio-economic conditions.
CHAPTER ONE: INTRODUCTION

1.1 Research background

Urbanization is a process by which a rural society is transformed into an urban one. Urbanization process may be driven by economic, demographic, political, cultural, technological, and environmental changes (Knox, 1994). Urbanization has been often accompanied by urban expansion and landscape changes. In past decades, rapid urban expansion and landscape changes have been observed in both developed and developing countries. Angel et al., (2005) found that the total built-up area of cities with population more than 100,000 in 1990 had increased by one-third during the period of 1990-2000, and that the growth had happened approximately equally between developing and industrialized countries. Seto et al. (2011) reported that India, China, and Africa had experienced the highest rates of urban land expansion, and that North America had made the largest change in total urban extent from 1970 to 2000. The proportion of urban population was 49% (3.2 billion) in 2005 while it was 13% in 1900 (United Nations, 2005), which has been an important factor driving the observed urban spatial expansion, particularly in the developing countries. In the post-industrial societies where population growth has more or less stabilized, rapid urban spatial growth has also occurred, primarily owing to suburbanization associated with the increasing popularity of automobiles and the post-industrial changes to economy (Clarke et al., 1997).

It is critical to monitor and model urban growth and landscape changes due to three reasons. Firstly, spatially continuous or discontinuous urban growth, when aggregated globally, can affect some key aspects of Earth System functioning (Lambin et al., 2001; Foley et al., 2005), such as global climate, hydrological systems, and ecosystems. In recent decades, globalization has gradually deepened. The removal of barriers separating world regions has enhanced the exchange of goods, technologies, population, and cultures (Wind, 1986). The magnitude and complexity of urban growth have substantially increased, particularly in the developing which cause many environmental problems, such as deforestation, desertification, soil erosion, biodiversity losses, climate change, and hydrological perturbations (Seto and Fragkias, 2005; Yu and Ng, 2007). Secondly, urban spatial growth and landscape changes can bring some negative
effects upon the urbanization process, which may prevent urban systems from sustainable
development. For instance, some issues closely associated with urban spatial expansion, such as
traffic congestion, air/water quality deterioration, poverty, crime, and unemployment, may affect
urban sustainable development. Thirdly, urban spatial growth and landscape changes are results
of the urbanization/suburbanization process as discussed above. Monitoring and modeling such
growth can provide useful insight into the urbanization process. In other words, measuring and
analyzing emergent urban patterns may be an effective way to understand how certain factors
interact over space and time and what functions they serve, and eventually to investigate urban
growth dynamics (e.g., Clarke et al., 1997; Dietzel et al., 2005; Seto and Fragkias, 2005; Hu and
Lo, 2007; Serra et al., 2008).

However, the complexity of urban systems challenges the study of urban growth and landscape
changes. As stated above, urbanization process can be driven by many forces, such as economic,
demographic, political, cultural, and biophysical factors. Their contributions to urban spatial
growth can vary over space and time. Urban growth monitoring and modeling can improve the
understanding of urban spatial and temporal dynamics. Because of the interaction with their local
periphery and the impacts of globalization, urban systems are becoming more dynamic and
complicated (Seta and Fragkias, 2005). Besides, it is hard to model the “feedback” effects of
urban expansion on the urbanization process. The time lag of these effects makes the simulation
even more complex (Liu et al., 2007). For instance, ecological degradation caused by urban
expansion may change human behavior and land use policies, but their causal relationship is hard
to simulate because of its nonlinearity and feedback effects. In addition, the sociality of cities
increases the difficulties of measuring urban development using equilibrium theory (Clarke et
al., 1997), such as influences of ethnics, religion, culture and life style on urban spatial growth.
Given the complexity of urbanization process, an approach integrating remote sensing, artificial
neural networks, and geographic information systems (GIS) can be useful to characterize urban
growth patterns and processes, which can help understand the spatial and temporal dynamics of
urban spatial growth.

Due to its technological robustness and cost-effectiveness, remote sensing has been increasingly
used to derive land cover information through either manual interpretation or automated
classification (Golubiewski and Hussein, 2007). The latter is more desirable but can be quite challenged when classifying heterogeneous, complex landscapes, such as urban land covers. Conventional statistic classifiers, such as maximum likelihood method, assume that the spectral signals for urban lands have a normal distribution with unknown mean and variance. The mean and variance are calculated from known training samples using maximum likelihood estimate. Urban surfaces are, however, very diverse. They can be concrete-made large roof buildings, glass buildings, wood-made houses, asphalt roads, paved large parking lots, or unpaved local roads. The diversity of urban surfaces makes the spectral signatures of urban lands have multiple clusters and obviously against the assumption of the normal distribution of data by conventional statistic classifiers. The existence of different percentages of vegetation covers (e.g., trees, small grasslands in urban lands) makes the feature spaces of urban lands even more complex. Therefore, advanced techniques are needed to improve land cover classification from remotely sensed data (Rogan et al., 2008). Based on the principles of the biological neural networks system, artificial neural networks (ANNs) are advanced intelligence techniques that are particularly suited to deal with nonlinear, complex phenomena (Haykin, 1999). They can simulate and present the multi-cluster feature spaces of urban lands and don’t have the assumption of the normal distribution of data. And they can also be effective to process vast quantities of data from various sources (Haykin, 1999). These advantages make neural networks an attractive pattern classifier for heterogeneous, fragmented landscapes in the urban environment. Over past two decades, many applications have proved that ANNs can outperform conventional statistical classifiers in terms of classification accuracy, such as Gaussian maximum likelihood (GML) (Benediktsson et al., 1990; Hepner et al., 1990; Bischof et al., 1992; Civco, 1993; Paola and Schowengerdt, 1995a; Serpico et al., 1996; Bruzzone et al., 1997). The integration of artificial neural networks and remote sensing can help improve land cover mapping from remote sensor data in a complex urban environment.

Besides remote sensing and artificial neural networks, GIS can be helpful for the study of urban growth and landscape changes. GIS-based spatial analysis and landscape metrics are useful to examine urban growth patterns based on urban land change information extracted from remotely sensed data by artificial neural networks. Furthermore, GIS-based spatial analysis can explicitly
demonstrate the spatio-temporal patterns of urban growth, and analyze urban growth processes based on the observed urban growth patterns.

Therefore, integrating artificial neural networks with remote sensing and GIS can be useful for improving urban land change detection, characterizing urban spatial growth, and providing insights into urbanization mechanics. The integration of these three technologies forms an invaluable technological framework that can help examine urban growth dynamics through the study of urban forms, functions, and their interrelations.

1.2 Research objectives

Given the importance and challenges in urban growth and landscape change studies, the overall aim of this dissertation is to improve our understanding of the spatial and temporal dynamics of urban spatial growth through the integration of artificial intelligence, image analysis and GIS from a perspective of the complexity theory. The dissertation will explore how artificial neural networks, remote sensing, and GIS can be combined to detect urban land changes, characterize urban spatial growth patterns and processes, and investigate urban growth dynamics at different scales in a rapidly urbanizing environment. The specific objectives include:

1. To improve land cover classification from remote sensor data by using artificial neural networks;
2. To characterize urban spatial growth patterns and processes by integrating artificial neural networks, remote sensing, and GIS technologies; and
3. To understand spatial and temporal dynamics of urban growth in a rapidly urbanizing large metropolitan area.

These objectives and goals will be addressed through imagery-based urban spatial growth characterization. Central to this research effort is the development of a technological framework that tightly integrates satellite imagery processing, artificial intelligence, and GIS-based pattern analysis. Specifically, the project will use remote sensor data and ANNs to extract urban land change information in a rapidly growing metropolitan area, and then analyze the spatial growth patterns at different scales by using GIS-based landscape metrics. Furthermore, the project will examine the urban growth processes based on the observed patterns. Finally, it will explore
spatial and temporal features of urban growth based on urban growth patterns and processes, which can help improve our understanding of urban growth dynamics. This dissertation project contrasts with past works on urban land changes that have focused primarily on urban land change detection from remote sensor data without providing in-depth insight into urban spatial growth patterns and the relevant urban growth processes.

1.3 Selection of the study area

This dissertation project will choose the Beijing metropolitan area as the case study site. Three major reasons lead to this choice. First, Beijing is one of few metropolises, which are stepping into post-industrial societies under the effect of globalization in the developing countries (Feng et al., 2008). Studies on Beijing have revealed a mode of urban development which jumps or compresses one or more historical phases in the evolution of urban space and becomes a complex urban system with elements of both traditional and modern societies (Gaubatz, 1995). These research efforts primarily examined the social-economic reforms, urban function transformation, and urban space development, but with little attentions on urban spatial growth patterns and processes that may be different from those in the traditional post-industrial cities. My research will explore how Beijing’s urban form changes in the context of globalization and the “info” revolution.

Secondly, Beijing is a commercial, industrial, cultural and political center in China, which has experienced a rapid growth in the past decades. The build-up area in Beijing has increased from 346 km$^2$ to 490 km$^2$ during the period of 1980-1999 (Xu et al., 2002). The dramatic changes have occurred not only in its built-up area, but also in urban form, urban functions and urban lives at every level since the late 20th century (Gaubatz, 1999). The transformation of urban form and structure in this city has been a consequence of social, economic, and political changes. The complexity of urban development makes this city an excellent site to examine the utilities of the methodology proposed in this dissertation.

In addition, due to its long history, landscapes in the Beijing metropolitan area are extremely heterogeneous. Previous urban landscapes, which were gradually formed by the city’s historical functions and productivity, have to be reconstructed, reserved, or transformed to meet the needs
of sustainable socio-economic development and modern urban lives, as well as historical culture protection. Therefore, comparing to the newly developed contemporary cities, Beijing, as a city with a thousand-of- year history, has extremely complex landscapes and unique urbanization process. These features make Beijing a good choice to conduct land cover classification by artificial neural networks from remotely sensed data and to validate the effectiveness of the proposed methodology in such a complex environment.

1.4 Research approach and hypotheses

The study of urban spatial growth has been traditionally undertaken by using various analytical approaches. These approaches attempt to uncover universal laws and the ways in which they produce observable geographic patterns of cities. They are often grounded on economic theories (Angelsen and Kaimowitz, 1999). For instance, “Gravity model” predicts the development of cities in a region by examining flows of people, capital, ideas and commodities. Places in a shorter distance have a greater attraction. “Bid-rent theory” explores urban form by computing land rent and possible returns of various land uses (e.g., residential, retail and manufacturing) at a certain location. Concentric zone model describes an urban form that may be produced based on the Bid-rent theory. Harris-Ullman’s multiple nuclei model describes an urban form where multiple nuclei are gradually developed at some locations due to the popularity of modern transportation and the process of agglomeration. These nuclei become centers of urban future expansion. Nevertheless, these analytical approaches are usually based on some overly simplistic assumptions of a flat and uniform space that does not necessarily reflect the reality (Clarke et al., 1997).

Qualitative approaches have also been used for urban pattern analysis. They characterize urban form through observations and descriptive statistics. For example, “Edge City” by Garreau (1991) describes the distribution and concentration of commercial activities following the residence in suburb areas.

Some scientists have also attempted to utilize empirical approaches to model the relationships between urban form and underlying urban growth forces using statistical methods, such as regression models (e.g., Hu and Lo, 2007). These approaches quantify empirical relationships
from the practical data, but they are less effective to analyze urban future development under alternative management schemes.

To examine the spatial and temporal urban growth dynamics, this dissertation adopts an approach focusing on urban pattern-process analysis, with particular attentions paid upon the spatial and temporal variations and scale dependency of urban growth (Figure 1.1). In recognition of the complexity of urban system, the approach emphasizes the use of the complexity theory and urban theories as the conceptual framework to generate improved capabilities in support of urban land change detection and urban growth characterization. Technologically, this approach proposes a methodology with three essential components, namely, image processing, artificial neural networks, and GIS. These components are combined to build a comprehensive technological framework for urban growth research. Given the heterogeneous landscapes and cross-scale interactions of socio-economic factors over space and time, this technical integration of ANNs, image analysis and GIS allows acquisition and integration of sufficient accurate spatial data, multi-scale analysis of urban form, and characterization of urban growth patterns and processes.

Specifically, this dissertation will detect urban land change, characterize urban spatial pattern, and identify urban growth features with the consideration of four common phenomena related to complex systems: spatial heterogeneity, temporal heterogeneity, scale multiplicity, and non-linearity. The research will include two parts. One is the examination of artificial neural networks in land cover classification from remote sensor data. And the other is the urban growth characterization at the metropolitan, functional zone, and cell levels by using three approaches: urban land change geo-visualization and statistical descriptions, landscape metrics analysis, and moving windows analysis.

Several hypotheses in relation to the three objectives will be tested:
Figure 1.1  Conceptual and technological integration of the pattern-process analysis in support of urban growth research.
(1) Appropriately parameterized ANNs can help improve the performance of land cover classification for heterogeneous landscapes in an urban environment;

(2) The spatial and temporal variations of urban spatial growth can be characterized by integrating remote sensing, ANNs, and GIS technologies;

(3) Scaling significantly influences the measurement of urban growth patterns. It is critical to characterize urban spatial growth patterns and processes at different scales;

(4) The dynamics of urban spatial growth can be examined by characterizing urban growth patterns and processes;

(5) Urban land changes in the Beijing metropolitan area can be effectively detected from remotely sensed data by ANNs; and

(6) The Beijing metropolitan area has experienced a rapid spatial growth, and the urban growth patterns are significantly different from most of the U.S. large cities.

1.5 Organization of the dissertation

This dissertation is organized into six chapters. Chapter One introduces the research background, research objectives, and the significance of the study, explains the selection of the Beijing metropolitan area as the case study, addresses the conceptual and technological frameworks of the research, and formulates research hypotheses. Chapter Two provides a literature review on some key techniques and methods used in this dissertation, including the complexity theory, remote sensing, artificial neural networks, and GIS-based spatial pattern analysis. Chapter Three describes the methodology used in this research. Specifically, it introduces the Beijing metropolis in terms of biophysical and socio-economic conditions and data collection and processing. Then, the Chapter focuses on the technical aspects of integrating artificial neural networks, image processing, and urban growth patterns and processes analysis. Chapter Four presents the procedures and experimental results of investigating the use of artificial neural networks for improving the performance of land cover classification from remote sensor data. Chapter Five addresses three approaches for urban growth characterization, including urban land
change detection and visualization; urban growth characterization by using the landscape metrics analysis; and urban growth dynamics analysis by using the moving windows analysis and GIS-based spatial analysis. This Chapter also presents major findings on urban growth features in the Beijing metropolitan area. Finally, Chapter Six summarizes the development and implementation of the concepts and technologies proposed in this dissertation, concludes the conceptual and methodological implications and limitations of this research based upon the case study in the Beijing metropolitan area, and discusses future research agendas.
CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

To address the research objectives discussed in Chapter One, this dissertation is concerned with several different theories and techniques. And this chapter reviews some core concepts and techniques for urban applications, which include complexity theory, remote sensing, artificial neural networks, and geographic information systems (GIS).

2.2 Complexity theory

Complexity theory refers to the study of complex systems. It is driven primarily by approaches from physics and mathematics and defines complex structure and behaviors (Cadenasso et al., 2006). At the epistemological level this theory sees the world from a different perspective, and at the methodological level it provides different methods and techniques to develop knowledge (Corpataux and Crevoisier, 2007). There are three major currents of complexity theory: algorithmic complexity from the perspective of mathematic and information theory, deterministic complexity through nonlinear analysis, chaos theory, and catastrophe theory, and aggregate complexity focusing on interactions of individual elements (Manson, 2008). The complexity research intends to investigate and model the complex structure/behaviors that primarily include reciprocal feedback loops, time lags, resilience, non-linearity, self-organization, path-dependence and emergence (An et al., 2005; Batty, 2005; Cadenasso et al., 2006; Liu et al., 2007; Manson, 2008).

An urban system is comprised of many subsystems and components such as urban functions production, residence, entertainment, transportation), human and environment (e.g., socio-economic development, water supply and demand, pollutions), and urban ecosystem (e.g., biodiversity, hydrological cycle, water and soil conservation). The interactions of these components are characterized by spatial and temporal variations and cross-scale which determines the complexity of urban systems (Wu and David, 2002; Cadenasso et al., Manson, 2008). This complexity inevitably increases the uncertainties of urban transformation and the difficulties for urban growth simulation and forecasting (An et al., 2005; Manson, 2006).
It is hard to model urban growth dynamics using traditional analytical methods (Batty, 1997; Almeida et al., 2008). Studying urban form, functions, and their interrelations should lead to a better understanding of urban spatial growth through a conceptual framework based on the complexity theory.

2.3 Remote sensing

Remote sensing measures and monitors important biophysical characteristics and human activities on Earth without physical contact through the use of electromagnetic radiation (Jensen, 2000). When compared to other data acquisition methods (e.g., in-situ measurement), remote sensing has many advantages. An important advantage is that remote sensing can collect data over a large geographic area at frequent temporal intervals without a sampling bias. In addition, remote sensing can provide some information on Earth that other methods cannot or can obtain only with great difficulty and/or cost. Especially in recent decades, Earth observation sensors are acquiring remotely sensed data in largely increasing quantities. These data become important resources for studying many environmental issues, including land use/cover change.

For urban applications, due to the broad coverage, frequent repetition and cost-effectiveness, medium-resolution remote sensor data are often used, such as Landsat, SPOT (Satellite Pour l'Observation de la Terre), and ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer). Among these remote sensor data, Landsat data are particularly useful for the study of urban land use/cover change because they have been available since 1970s.

Two types of techniques can be used to detect land use/cover change from remote sensor data: image-to-image comparison and map-to-map comparison. The former involves the comparison of two images acquired at different dates. It requires a radiometric calibration procedure because changes in spectral intensity caused by atmospheric conditions or sensors can bring large errors into results of land cover change detection. In addition, this method highlights changes between two images, but does not provide from-to information. The map-to-map comparison, also known as post-classification change detection, measures the categorical land cover changes occurring between a suit of thematic land-cover classes (Jensen, 2004). Image-image registration is needed by this method. Each remotely sensed image has to be classified and the resultant land use/cover
maps are then compared pixel by pixel. With map-to-map change detection method, since errors in each individual map are accumulated in the final land change map, it is imperative to make image classification as accurate as possible. The classification accuracy by conventional methods is challenged by landscape complexity. Advanced classification method is often needed in order to achieve satisfactory classification accuracy. In this dissertation research, I will use the map-to-map comparison method to detect urban land cover changes, and I will use artificial neural networks to map land cover information from Landsat imagery acquired at each individual date.

2.4 Artificial neural networks

A neural network is a massively parallel distributed processor comprised of simple processing units, attempting to simulate the powerful capabilities for knowledge acquisition, synthesis, and problem solving of the human brain (Haykin, 1999). It originated from the concept of artificial neuron introduced by McCulloch and Pitts in 1943. Over past several decades, neural networks have evolved from the preliminary development of artificial neuron, through the rediscovery and popularization of the back-propagation training algorithm, to the implementation of neural networks using dedicated hardware (Dawson and Wilby, 2001). Because of the distributed structure and adaptive learning process, neural networks are capable of handling non-linear, complex phenomena; and they can also be effective to process incomplete, noisy and ambiguous data (Bishop, 1995; Duda et al., 2001). These advantages make ANNs an accepted alternative to traditional statistical methods for improving classification accuracy.

The use of neural networks in remote sensing began in late 1980s (Atkinson and Tatnall, 1997; Kanellopoulos and Wilkinson, 1997). In past two decades, numerous studies have demonstrated that neural networks can produce identical or improved classification accuracies when compared to outcomes from conventional classifiers (e.g., Benediktsson et al., 1990; Bischof et al., 1992; Civco, 1993; Paola and Schowengerdt, 1995a; Gopal and Woodcock, 1996; Serpico et al., 1996; Mannan et al., 1998; Ji, 2000; Seto and Liu, 2003; Del Frate et al., 2007; Ashish et al., 2009; Petropoulos et al., 2010). The use of artificial neural networks is, however, challenged by the difficulty of network design and parameterization. Many factors affect the performance of including network topology, training algorithm and parameters setting, and network architecture (Paola and Schowengerdt, 1997; Duda et al., 2001; Ozkan and Erbek, 2003; Kavzoglu and
Mather, 2003; Zhou and Yang, 2010; Zhou and Yang, 2011). With the incorporation of neural networks as a standard classifier in some popular image processing software packages (cf. Mas and Flores, 2008), handling these diverse parameters present a challenge to beginners and even some experienced users as an inappropriate treatment can lead to suboptimal or unacceptable classification performance (Yang and Zhou, 2011).

Investigating the sensitivity of neural networks with respect to various parameter settings has been the subject in an increasing number of studies since this knowledge is critical to the design of efficient neural network models for improved performance (e.g., Korczak and Hammadimesmoudi, 1994; Yoshida and Omatu, 1994; Jarvis and Stuart, 1996; Kanellopoulos and Wilkinson, 1997; Ozkan and Erbek, 2003; Stathakis, 2009). These studies have been conducted by using a trial-and-error approach (e.g., Paola and Schowengerdt, 1997; Kavzoglu and Mather, 2003) or some advanced methods such as generic algorithm and pruning algorithms (e.g., Kavzoglu and Mather, 1999; Benediktsson and Sveinsson, 2003). As a result, some practical guidelines have been proposed to deal with various non-algorithmic issues including input data dimensionality and training sample quantity and quality (e.g., Zhuang et al., 1994; Foody et al., 1995; Kanellopoulos and Wilkinson, 1997; Mas and Flores, 2008). Nevertheless, there is no consistent guidance to help configure neural networks. Specific to the multi-layer-perceptron (MLP) neural networks, for example, there is no consensus on the number of hidden layers, type of activation functions, or training parameters that should be used to achieve optimal performance (Jain et al., 1996; Paola and Schowengerdt, 1997; Kavzoglu and Mather, 2003; Mas and Flores, 2008). These inconsistencies justify further research on the algorithmic issues in order to promote the routine use of neural networks in image classification.

In addition, adaptive training is critical for image classification by artificial neural networks (ANNs). Previous studies have shown that training algorithms can affect the network performance for some other applications, such as computational material science (Skinner and Broughton, 1995), face/non-face recognition (Tivive and Bouzerdoum, 2005), coplanar waveguides (Guney et al., 2006), thermal/pressure food processing (Torrecilla et al., 2007), and streamflow forecast (Kisi, 2007). However, little research has been directed to evaluate how various training algorithms could affect the performance of image classification by neural
networks. On the other hand, recent advances in computer software engineering allow incorporating neural networks as part of routine pattern recognition methods into different image processing software packages. While more network architectures being added into these software packages, virtually no practical guidelines have been developed on the use of training algorithms for image classification. Also, few studies have evaluated the performance of different neural networks architectures in image classification.

From the above discussion, it is clear that there are both advantages and limitations of applying artificial neural networks to image classification. On one hand, as an advanced intelligence technique, artificial neural networks can lead to an increase of land cover classification accuracy from remotely sensed data, and thus provide useful land cover information for urban landscape studies. On the other hand, integrating artificial neural networks with remote sensing and GIS has been challenged by the difficulties of appropriately parameterizing neural networks. Further research is needed to optimize the design of neural networks for the improvement of the performance of land cover mapping.

2.5 Geographic information systems

GIS can be thought of as theories and techniques designed to collect, store, manipulate, analyze, manage, and present geographically spatial data from the real world (Burrough, 1986). It is usually considered to have four components: geo-database; geo-spatial analysis; geo-modeling; and geo-visualization. GIS-based analysis can effectively derive useful information from large quantities of spatial data.

In urban studies, conventional GIS-based spatial analysis can examine spatial patterns of urban growth through spatial data integration and synthesis, and GIS visualization. They can measure the amount of land transformation at either landscape or class level. Yang (2002) quantified the land cover conversion between different classes using GIS analysis. Geo-spatial analysis can also investigate the diffusion or concentration of land cover changes at specific locations, such as urban centers, sub urban centers, and major transportation facilities (e.g., Kumar et al., 2007). But these techniques may be less effective for quantifying the spatial configuration of urban land changes.
Different from conventional geo-spatial analysis, landscape metrics analysis is particularly designed to quantitatively measure spatial patterns of geographic phenomena. In this dissertation research, landscape metrics analysis will be an essential component for the pattern-process analysis in order to explore urban spatial growth dynamics.

Landscape metrics were originally developed to measure landscape pattern that has been linked to ecological processes and from which the processes can be predicted (Burel and Baudry, 2003). They can characterize landscape properties at three levels: landscape metrics for entire patch mosaic comprised of different land classes, class metrics for a particular land class in the landscape, and patch metrics for an individual patch. A large group of landscape metrics have been developed to quantify the following aspects:

1) Landscape composition. These metrics quantify landscape composition, such as proportions, relative richness, and diversity. They are usually landscape-based metrics.

2) Patch properties. The metrics examine properties of patches, including patch number, size, shape, and density. They can be patch-based metrics. When aggregated, they can also be class-based or landscape-based metrics. While some properties are straightforward (such as patch number, size, and density), it can be complicated to measure some patch properties (such as patch shape). Shape complexity is an important indicator of landscape complexity. The simplest metrics for measuring shape complexity is the Perimeter-to-Area Ratio (PAR). This metrics is, however, sensitive to the patch size. In other words, the change of PAR can be caused by the changes in either shape complexity or patch size. Some metrics can measure the shape complexity without the size problem, such as Landscape Shape Index (the patch perimeter / the minimum possible perimeter with the same area).

3) Spatial configuration. The metrics evaluate the adjacency of landscape classes (degrees of aggregation or dispersion). They can be class-based or landscape-based metrics. Many metrics can measure landscape spatial configuration. Aggregation Index (AI) is a typical metrics for measuring the degree of aggregation, which equals to the number of like adjacencies between pixels of the interest class divided by the
maximum possible number of like adjacencies. Euclidean Nearest-Neighbor distance (ENN) reports the degree of isolation by calculating the distance of a patch to its nearest neighboring patch.

Besides ecological landscape, landscape metrics were found to be useful to characterize and quantify urban spatial patterns in recent years. Jenerette and Wu (2001) found that landscape metrics, such as patch density, edge density, fractal dimension, and contagion, were useful to capture the land use trends associated with urbanization in the central Arizona-Phoenix region. Herold et al. (2003a) applied landscape metrics to the analysis of spatio-temporal dynamics of urban growth in Santa Barbara, California. Seto and Fragkias (2005) showed that landscape metrics can help develop a comprehensive understanding of the shapes and trajectories of urban expansion in four Chinese cities. Consequently, landscape metrics become widely used in urban spatial pattern studies (Luck and Wu, 2002; Wu et al., 2006; DiBari, 2007; Aguilera, et al., 2011; Malinverni, 2011; Shrestha et al., 2012).

However, there are many landscape metrics available for pattern analysis, and some are highly correlated (Riitters et al., 1995; Griffith et al., 2000). It is important to carefully select a set of metrics to effectively measure landscape patterns, avoiding high correlations. Besides the dependency, as I stated above, landscape metrics were originally developed in landscape ecology to study the relationship between landscape patterns and ecological processes. Due to the differences in urbanization processes and ecological processes, only some landscape metrics can be useful to explore urban patterns, urban processes, and their relevance. So it is crucial to select appropriate ones from a large number of metrics.

Scaling is another issue in the use of landscape metrics. The cell size may influence the recognition of urban land change patches, and thus affect the results of spatial pattern analysis. Also, the modifiable areal unit problem (MAUP) may affect the performance of landscape pattern analysis, such as different zonings. In brief, the scale multiplicity of urban growth challenges the use of landscape metrics for urban growth characterization (Yu and Ng, 2007).

In addition, few landscape metrics studies have been done to investigate the implications of landscape metrics for urbanization processes and urban future development (Li and Wu, 2004).
Urban spatial configuration brings feedback effects on urban sustainable development while urban form is shaped by natural and socio-economic processes. Thus, it is feasible to investigate the underlying urban growth processes from observed urban patterns through a pattern-process analysis.
CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This dissertation research aims to characterize urban spatial growth in a rapidly developing urban environment based on a conceptual framework drawn from the complexity theory and a technological framework that integrates image processing, artificial neural networks, and GIS technologies. There are different currents of complexity theory, addressing the complex phenomena from different perspectives, as stated in Section 2.2. The complexity identified here refers to the phenomena that system components interact over space and time in a complex, non-linear fashion. The complexity theory considers a system as whole that can’t break into separate independent parts and that is not a perfect aggregation of individual components (Kocabas and Dragicevic, 2006). Therefore, the technique of artificial neural networks, which is particularly useful to deal with non-linear complex relationship, is used to improve urban land change detection from satellite imagery. And a pattern-process based urban growth characterization method is developed to explore the spatial and temporal features of urban growth.

Specifically, an improved methodology is proposed to achieve the research goals formulated in Chapter One. As illustrated in Figure 3.1, the research methodology used in this doctoral dissertation research project includes five major working phases: (1) selection of the case study site; (2) data acquisition and collection; (3) urban land change mapping from satellite imagery; (4) urban spatial growth characterization; and (5) result presentation and documentation.

In phases 3 and 4, I first examine the use of artificial neural networks to derive urban land changes from satellite imagery. Next, I integrate different sources of spatial data in a GIS environment and quantify urban spatial growth patterns using descriptive statistics, landscape metrics analysis, and moving windows analysis at different scales. Finally, I summarize the urban land development through the exploration of the spatial and temporal properties of urban growth patterns and processes. The detailed working procedural route is illustrated in Figure 3.2.
Figure 3.1  The major working phases in the dissertation research

- Phase One: Selection of the Case Study Site
- Phase Two: Data Acquisition & Collection
- Phase Three: Urban Land Change Mapping from Satellite Imagery
- Phase Four: Urban Spatial Growth Characterization
- Phase Five: Result Presentation & Documentation
Figure 3.2 Working procedural route.
3.2 Study site

I have discussed the rationale behind the choice of the Beijing metropolitan area as my study area in Chapter One. Here I will give an overview of the geography, physical environment, history, population growth, economy, and urbanization process in the study area.

Geography and physical environment

The geographic coverage of the Beijing metropolitan area consists of the four functional zones, with a total area of approximately 16410.54 km², fitting within two Landsat TM scenes (185*185km for each scene). The four functional zones, which were defined in the “Beijing regional development planning at the Eleventh Five-Year period” published by the Beijing Municipal Commission of Development and Reform, are composed of 18 districts/counties and differentiated by administrative boundaries (Figure 3.3):

1) The urban core zone, including districts Dongcheng, Xicheng, Xuanwu, and Chongwen;
2) The extensive urban zone, including districts Haidian, Chaoyang, Shijinshan, and Fengtai;
3) The new urban zone, including districts Changpin, Shunyi, Tongzhou, Fangshan, and Daxin; and
4) The ecological conservation zone, including counties Yanqing, Miyun, Pingu, Huairou, and Mentougou.

Physiographically, Beijing is located at the northern edge of the North China Plain (Latitude: N 39°C; Longitude: E 116°C) (Figure 3.3). It is composed of the alluvial plains of Rivers Yongding and Chaobai in the south and east and hills/mountains in the north, northwest, and west. The plains have an average altitude of 20-60 m, and the mountains are generally 1000 to 1500m above the sea level with the highest peak of 2303m. Five major river systems run through the area from the west to east, which include Rivers Juma, Yongding, North Canal, Chaobai, and Jiyun. In addition, there are 82 reservoirs, of which large ones include Miyun, Guanting, Huairou, and Haizi. The climate is characterized by hot, humid summers and generally cold, windy, dry
Figure 3.3  Location of the case study site. It consists of the four functional zones, namely, urban core, extensive urban, new urban, and ecological conservation region.
winters. The annual average temperature is 13.3 Celsius degree. Approximately 75% of the annual rainfall is concentrated in the summer months (Beijing government, 2008).

Due to the complex terrain and ecological diversity, the Beijing metropolitan area is rich in vegetation species and marked with vertical vegetation distribution. However, human activities in the past hundreds of years have a profound impact on the landscape. According to historical records, Beijing area was covered by dense forests a thousand years ago. After that, population growth, farming, urbanization and wars destroyed most of the ancient forest. Forest coverage was only 1.3% in 1949, and gradually increased to 10.32% by 1980 (Li, 2008). At present, the plains (38.6% of total area) are dominated by farmland and towns, and the mountainous areas (61.4% of total area) are by secondary broad-leaved, deciduous shrubs with small quantities of deciduous and coniferous forests. The landscape complexity in the study area makes conventional methods less effective for urban land change detection from remote sensor data. Artificial neural networks are used as an acceptable alternative to improve image classification.

**History**

Beijing is the capital of the People’s Republic of China (PRC) and one of the fastest growing cities in the world. It is also an ancient city with a rich history. It served as a capital of China for five dynasties for 800 years, namely, Liao, Jin, Yuan, Ming, and Qing. The long history of the city gives it unique urban form and development trajectory.

**Population**

The population in the Beijing metropolitan area has experienced a rapid growth, particularly during the past decade (Figure 3.4). The total population was 17.55 million in 2009 with 14.92 million urban populations while it was 13.57 million in total in 2000. The population explosion inevitably caused dramatic changes in urban form, such as urban expansion.

**Economy and urbanization**

Beijing is one of the most important economic centers in China, and has had great economic growth in the past four decades. According to the Census data by the National Bureau of Statistics of China, the total GDP (gross domestic product) in the Beijing metropolitan area was
Figure 3.4  The population growth in the Beijing metropolitan area. (a) Populations from 1960 to 2009; and (b) populations between 2000 and 2009. (Source: China’s National Bureau of Statistics)
204 billion US dollars in 2010, which was 31.5 times greater than that in 1978 in terms of the absolute values. Considering the purchase power, the average annual GDP growth was 10.4% from 1979 to 2007, which was higher than the national average annual growth (9.8%). And the 2008 Olympic game had contributed to the extremely high GDP growth in Beijing after 1999 (Figure 3.5).

Since the economic reform in 1978, the planning economy has been gradually transformed into the state intervention market economy. Beijing’s economic structure has then changed tremendously (Figure 3.6). Firstly, the status and ratios of the primary industry and the secondary industry have decreased while the tertiary industry (i.e., the service industry) with information-technology as the core component has developed rapidly. Secondly, the changes of industrial structures have brought changes on the occupation structures. Numerous labor forces have entered into the service industry and administration functional divisions. In 2007, Beijing became the most advanced area in terms of the tertiary industry and its GDP of the tertiary industry accounted for 71.35% of the total GDP, ranking No.1 among all provinces or municipalities under the jurisdiction of the State Council in China (source: China’s National Bureau of Statistics). In recent years, the real estate industry and automotive industry in Beijing have experienced a rapid development. By December 2009, the total number of vehicles in Beijing exceeded 400 million (Xinhua news, 2009).

In early 2005, the State Council of China officially approved the “Beijing City Comprehensive Plan (2004-2020)”. This plan re-defines the nature of the city of Beijing as the capital of China, the state’s political and cultural center, and the world-renowned ancient capital and modern international city.

To sum up, some dramatic changes have occurred in the Beijing metropolitan area in terms of population, economy, and land use policies. These changes have inevitably affected the development of urban spatial structure, which in return promotes or constrains the future population and economy growth. Monitoring and modeling urban spatial growth in the Beijing metropolitan area can provide insights into the mechanics of urban spatial development in a rapidly growing area with complex socio-economic conditions. This dissertation will take a close
Figure 3.5  Trends of Gross Domestic Product (GDP) in the Beijing metropolitan area. (a) GDP growth from 1978 to 2010; and (b) GDP annual growth rates from 1999 to 2007. (Source: Beijing’s Statistics Yearbook)
Figure 3.6  Industrial structure evolution in the Beijing metropolitan area from 1978 to 2008. Note that blue is for the primary industry; red for the secondary industry; and green for the tertiary industry (service industry) with info-technology. (Source: Beijing’s Statistics Yearbook)
look at the urban growth dynamics by integrating remote sensing, artificial neural networks, and GIS technologies.

3.3 Data acquisition and collection

3.3.1 Satellite images

Due to the technological robustness, a variety of remote sensor data are available for different applications. The selection of appropriate satellite imagery relies on many factors, such as research objectives, budgets, image processing technologies, spatial and temporal resolutions, and time duration (Jensen and Cowen, 1999; Herold et al., 2003b). Project objectives to a large part determine the demands on image resolutions, the time span, the image coverage, and so on. Also, the selection of remote sensor data needs to consider project budgets and image processing technologies.

Since the research objectives here are to investigate urban spatio-temporal dynamics, both high-resolution and moderate resolution images should be considered (Jensen and Cowen, 1999; Ward et al., 2000). However, high-spatial resolution images (such as images from IKONOS and QUICKBIRD) are costly and demand more processing and computing resources. Moreover, the earliest high resolution images became available after 2000 while Landsat imagery can trace back to 1976. Landsat imagery is quite competitive for the temporal change analysis, particularly during a long time period. This dissertation chose to use two sets of Landsat images as the primary data for the study of urban growth dynamics. The period of 2000-2009 is targeted because of the following reasons. Urbanization has been accelerated by the economic development since China opened its door to the western world in 1978 (Yu and Ng, 2007). Especially after China was accepted into the World Trade Organization (WTO) in 2001, China’s economy recovered from the 1998 Asian financial crisis, and started growing at a very high rate. Over 10% annual increase in GDP at average was reported by China’s National Bureau of Statistics (see Figure 3.5). Beijing has experienced complicated urban growth and landscape changes due to the dramatic socio-economic development during this period (Feng et al., 2008). In addition, urban population has greatly increased since 2000 (see Figure 3.4). The time span of 2000-2009 is appropriate for investigating urban growth dynamics in a rapidly urbanizing area.
Because the study area spans two columns in the WRS (Worldwide Reference System) that is used for indexing locations of Landsat scenes, two scenes were mosaicked to cover the entire study area for each date. Table 3.1 summaries the specific dates, types of remote sensors, Landsat satellite series number, nominal spatial resolution of the various sensors, and other environmental parameters for each scene.

3.3.2 Reference data

Reference data play an important role in improving the performance of land cover classification from remote sensor data. They may be used in various aspects of image classification, including the classification scheme design, training sample collection, and classification results validation. They may also provide information for geographically interpretation of resultant urban growth patterns. Reference data used in this research include: (1) Beijing administrative boundary maps; (2) The 2007 land use/land cover map at the Haidian district produced by China’s Ministry of Land Resources; (3) SPOT satellite images acquired in May, June, and October, 2004, whose spatial solutions are 2.5m for panchromatic band and 10 m for multi-spectral bands; (4) Two QUICKBIRD images acquired on April 24, 2002, whose spatial resolutions are 60-70 cm for panchromatic band and 2.4-2.8m for multi-spectral bands; and (5) digital elevation model (DEM) data with 10 m resolution.

3.3.3 Field work

I have conducted GPS-guided field work twice in Beijing in 2008 and 2009, respectively. Purposes of the field work include: (1) obtaining the first-hand information to better understand the natural and cultural landscapes in Beijing; (2) establishing reliable connections between image signals and ground objects in support of land cover classification from satellite imagery; (3) collecting socio-economic data for the urban spatial growth analysis.

The first field work was conducted in may-June, 2008, and had targeted to the first two tasks. To gain a general understanding of the landscapes, I traveled across the entire study area, from the urban core to newly developed towns and villages, from mountains to plains, and from commercial areas to residential and agricultural areas. Figure 3.7 shows the routes of the ground-truth survey. I also collected field data for each major land use and land cover category for image
Table 3.1  List of satellite images used for land use/land cover mapping.

<table>
<thead>
<tr>
<th>Year</th>
<th>Acquisition Date</th>
<th>Landsat No.</th>
<th>Remote Sensor</th>
<th>Nominal Spatial Resolution (m)</th>
<th>Path/Rows</th>
<th>Mosaic*</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>April 30, 2000</td>
<td>7</td>
<td>ETM+</td>
<td>28.5×28.5**</td>
<td>32/123</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>April 30, 2000</td>
<td>7</td>
<td>ETM+</td>
<td>28.5×28.5**</td>
<td>33/123</td>
<td>Yes</td>
</tr>
<tr>
<td>2009</td>
<td>September 22, 2009</td>
<td>5</td>
<td>TM</td>
<td>28.5×28.5***</td>
<td>32/123</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>September 22, 2009</td>
<td>5</td>
<td>TM</td>
<td>28.5×28.5***</td>
<td>33/123</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* A mosaic operation was conducted to combine two scenes acquired at the same date in order to cover the entire study area.

** The panchromatic band has a spatial resolution of 15 m; and the thermal band is 60m.

*** The spatial resolution of the thermal band is 120 m.
Figure 3.7  Routes of ground-truth survey in the Beijing metropolitan area.
classification. For better positioning, I used a Trimble GPS receiver, along with Beijing topographic and road maps, to locate each observation site. Besides collecting geographic coordinates using a GPS receiver, I also took photos and field notes and linked them to the relevant GPS points. Figure 3.8 shows what ground features look like on satellite images by geocoding GPS points and relevant photos to satellite imagery. Through the integration of GPS data, photos, and images, I established connections between image signals and actual land covers. The second field work was completed in May-June, 2009, and I collected socio-economic data and supplementing ground data. I visited relevant organizations/departments (e.g., Bureau of Statistics) of the 18 districts in Beijing and requested statistical data. These data include economic data (e.g., gross industrial product, per capital income, per capital consumption, economic structure, and real-estate development), demographical data (e.g., number of households, population, agricultural population, and non-agricultural population), transportation (e.g., road millage), and social life (e.g., per capital house area, vehicle per household, and education). I also visited some critical locations, such as the urban core, suburban centers, industrial development zones, newly constructed highways, and residential areas, in order to understand the forces driving urban growth.

3.4 Urban land change mapping

This part of research work was designated to address the first research objective, as stated in Chapter One. In doing so, an experiment-based method was adopted to examine the use of artificial neural networks for image classification. The northern Atlanta metropolitan area, Georgia, USA was used as a case study site because it had experienced rapid urban growth in past decades, as what the Beijing metropolitan area did. The fragmented and heterogeneous landscape in this area made it an appropriate site for examining the effectiveness of artificial neural network in land cover classification in a complex urban environment. As a result, a systematic approach that can guide the use of neural networks for image classification from remote sensor data was proposed. Then, this approach was applied to urban land change mapping in the Beijing metropolitan area from multi-temporal remote sensor data.
Figure 3.8  Landsat TM images and relevant photos showing different land use/land cover types. Red points are GPS data collected at locations where photos were taken. Through geographic coordinates of these GPS data, the photos can be directly geo-coded to satellite images. (a) Single family houses; (b) Corn land; (c) Industrial development zone.
The design of the method was based on two assumptions. One is that artificial neural networks have the potential of improving the accuracy of land cover classification. This has been confirmed by many prior studies (e.g., Benediktsson et al., 1990; Bischof et al., 1992; Civco, 1993; Paola and Schowengerdt, 1995a; Gopal and Woodcock, 1996; Serpico et al., 1996; Mannan et al., 1998; Ji, 2000; Seto and Liu, 2003; Del Frate et al., 2007; Petropoulos et al., 2010). The other is that the performance of neural networks is contingent upon a wide range of algorithmic and non-algorithmic parameters, such as input data dimensionality, training data, network structure, and learning process. This issue has been brought up by many researchers (e.g., Paola and Schowengerdt, 1995b; Foody and Arora, 1997; Kavzoglu and Mather, 2003; Mas and Flores, 2008).

Specifically, to examine the use of artificial neural networks for improving image classification, intensive experiments were carried out to address three important issues. “How do internal factors affect image classification accuracy by multi-layer-perception (MLP) feed-forward neural networks?” “What are the performances of different training algorithms in image classification by MLP neural networks?” “What are the impacts of neural networks architectures on land cover classification?” These experiments provided insights into the use of artificial neural networks for urban land change mapping from remote sensor data. They will be discussed in details in Chapter Four.

3.5 Urban growth characterization

This part of research work was designated to investigate the spatio-temporal urban growth patterns, urban growth processes, and their relevance, which addressed the second and third objectives in this dissertation. In doing so, statistical analysis, landscape metrics analysis, and moving windows analysis were combined to characterize urban growth at the metropolitan, functional zone, and cell levels from different perspectives. Different from prior studies, this research concentrated on the pattern-process analysis and urban growth characterization without driving forces analysis and computationally intensive simulations. In other words, this research aimed at measuring urban spatial growth pattern, analyzing urban growth process, and discussing spatial and temporal features of urban growth from observed urban patterns and processes.
First of all, a method integrating artificial neural networks with a wide range of remote sensing and GIS technologies was used to detect and visualize urban land changes in the Beijing metropolitan area. It primarily involved the following procedures: image preprocessing, image classification by neural networks, image post-classification, urban land change detection, and urban growth pattern analysis. The detailed procedures will be discussed in Section 5.2.

Secondly, landscape metrics analysis was adopted to characterize urban growth patterns at the metropolitan and functional zone levels. Scale dependency of landscape metrics analysis was investigated from two aspects: cell size and map extent. Urban growth patterns were evaluated in terms of landscape fragmentation, landscape complexity, landscape isolation, and landscape aggregation. Three major urban growth processes were considered: urban compaction, urban aggregation, and urban dispersion. The methodology and major findings will be discussed in Section 5.3.

Finally, landscape metrics analysis with moving windows and GIS-based spatial analysis were applied to characterizing urban spatial growth pattern at the cell level and exploring urban growth dynamics based on pattern-process analysis. Moving window analysis was used to quantify the spatial variations of urban spatial growth. GIS-based spatial analyses were further utilized to reveal the spatial and temporal evolution of urban patterns and processes.
CHAPTER FOUR: ARTIFICIAL NEURAL NETWORKS FOR LAND CLASSIFICATION

4.1 Introduction

Land cover classification by artificial neural networks (ANNs) from remote sensor data involves in three primary procedures: image data acquisition and preprocessing; neural networks parameterization and training; and image classification and accuracy assessment. The most challenging part is to appropriately parameterize neural networks, which has been discussed in Chapter Two. This chapter will review and assess a set of algorithmic and non-algorithmic parameters affecting the performance of neural networks in land cover classification from remote sensor data. The chapter comprises several major components. Firstly, I introduce and review fundamental aspects of neural networks, including neural network architectures and knowledge representation. Secondly, I discuss three focused studies by which I assessed the sensitivity of neural networks with respect to various internal parameter settings, evaluated the performance of several training algorithms, and investigated the impacts of network architectures in image classification. Thirdly, based on literature review and my focused studies, I propose a framework to guide the use of neural networks in image classification, considering data acquisition and preprocessing, network model design, training process, and validation in a sequential mode.

4.2 Fundamentals of ANNs

This section will discuss neural network architectures with the emphasis upon the multi-layer perceptron (MLP) networks, along with neural networks training methods.

4.2.1 Neural network architectures

There are two fundamentally different neural network architectures: feed-forward networks and recurrent networks. The former includes single-layer networks comprising an input layer that projects onto an output layer, as well as multilayer networks having at least one hidden layer that allows the networks to extract high-order statistics. A recurrent network distinguishes itself from feed-forward networks by having at least one feedback loop whose presence can greatly affect the training capability and performance (Haykin, 1999).
Considering neural network structures and training paradigms, we can find a large number of different types of neural networks, and some of them are listed in Table 4.1. Each type has advantages and limitations depending upon applications. Detailed discussions about these neural network types are given elsewhere (e.g., Bishop, 1995; Jain et al., 1996; Rojas, 1996; Haykin, 1999; Principe et al., 2000). I particularly introduce four commonly-used neural networks, namely, multi-layer-perceptron (MLP) feed-forward networks, Kohonen’s self-organizing-map (SOM) networks, adaptive-resonance-theory (ART) networks, and probabilistic networks. MLP neural networks and ART neural networks are chosen to evaluate the performances of two basic neural network structures: the simplest feed-forward structure and the recurrent structure. SOM neural networks are chosen because of their popularity and their unique neighborhood-based competitive training process (Duda et al., 2001). And probabilistic neural networks are chosen because of its extreme ease of implementation and training when compared to other neural networks (Zhang et al., 2009).

4.2.1.1 MLP neural networks

The MLP neural networks are relatively easy to understand and implement. As the workhorse of neural networks, they have been increasingly used in remote sensing (cf., Mas and Flores, 2008). They comprise distributed neurons and weighted links (Figure 4.1). Arranged in an input-hidden-output layered structure, each neuron contains a simple processing function (i.e., activation function) that individually handles pieces of complex problems; the weighted links between neurons determine the direction of data flow and the contribution of the “from” neuron to the “to” neuron. These weights can be determined through an iterative back-propagation training process that learns from known samples and adjusts the weights between neurons until the minimum error of the performance function is achieved. While the MLP structure and the concept of back-propagation algorithm are relatively simple, network topology and training parameter settings can complicate the overall performance of neural networks for image classification (cf., Foody and Arora, 1997; Kavzoglu and Mather, 2003; Mas and Flores, 2008).

4.2.1.2 SOM neural networks

SOM neural networks, also called Kohonen’s SOM networks, are techniques of representing multi-dimensional data using fewer dimensions but preserving topological information as much
Table 4.1 List of some neural networks in machine learning.

<table>
<thead>
<tr>
<th>No.</th>
<th>Network Types</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multi-layer feed-forward perceptron networks</td>
<td>They comprise an input-hidden-output layered structure, and the input signal propagates through the network in a forward direction. They are the most widely used networks.</td>
</tr>
<tr>
<td>2</td>
<td>Radial basis function networks</td>
<td>They use radial basis functions to replace the sigmoidal hidden layer transfer function in multi-layer perceptrons, thus transferring the design of a neural network as a curve-fitting problem in a high-dimensional space.</td>
</tr>
<tr>
<td>3</td>
<td>Self-organizing networks</td>
<td>They use a supervised or an unsupervised learning method to transform an input signal pattern of arbitrary dimension into a lower dimensional (usually one or two dimensional) space with topological information preserved as much as possible.</td>
</tr>
<tr>
<td>4</td>
<td>Probabilistic networks</td>
<td>They use a non-linear activation function to transform inputs, and output neurons sum the contributions from all neurons connected to them. They have high space complexity, but very efficient in training and pattern recognition.</td>
</tr>
<tr>
<td>5</td>
<td>Recurrent network (e.g., Hopfield network; Adaptive resonance theory network)</td>
<td>Contrary to feed-forward networks, recurrent neural networks use bi-directional data flow and propagate data from later processing stages to earlier stages.</td>
</tr>
<tr>
<td>6</td>
<td>Modular neural networks (e.g., Committee of machine)</td>
<td>They use several small networks that cooperate or compete to solve problems.</td>
</tr>
<tr>
<td>7</td>
<td>Stochastic neural networks (e.g., Boltzmann machine)</td>
<td>This type of networks introduces random variations, often viewed as a form of statistical sampling, into the networks.</td>
</tr>
<tr>
<td>8</td>
<td>Neuro-fuzzy networks</td>
<td>They are a fuzzy inference system in the body which introduces the processes such as fuzzification, inference, aggregation and defuzzification into a neural network.</td>
</tr>
</tbody>
</table>

* Detailed discussions on these neural network types are given elsewhere (e.g., Bishop, 1995; Haykin, 1999; Duda et al., 2001).
Figure 4.1 A fully-connected multilayer perceptron (MLP) neural network with a 4 X 5 X 4 X 2 structure. This is a feed-forward architecture. Data flow starts from the neurons in the input layer and moves along weighted links to neurons in the hidden layers for processing. Each hidden or output neuron contains a nonlinear discriminant function that combines information from all neurons in the preceding layers. The output layer is a complex function of inputs and internal network transformations (Yang and Zhou, 2011).
as possible (Kohonen, 1982; Ji, 2000; Duda et al., 2001). They are capable of visualizing feature spaces by data dimension reduction to as few as two. Like other neural networks, SOM networks are particularly useful when desired patterns are in a nonlinear fashion with the data. Kohonen’s SOM neural networks consist of an input layer receiving the input vector (e.g., image bands) and an output layer fully connecting input neurons via weighted links (Figure 4.2). These weights have to be optimized through an iterative training process. During the training process, each output neuron computes its net activation that reflects the similarity between this neuron and the input training sample. Net activation can be computed by:

$$net_k = x^t w_k$$  \hspace{1cm} (1)

where $x^t$ refers to the transpose of input vector, and $w_k$ is the weight vector between the output neuron $k$ and input neurons. The output neuron with the largest net activation is the *winner* and becomes the center of a recognized pattern (Richardson et al., 2003). Weights of the links to the winner and its neighbors within a user-defined radius are then updated by the following equation:

$$w_{kl}(t + 1) = w_{kl}(t) + \eta(t) \Lambda(|y - y^*|)(X_l - w_{kl}(t))$$  \hspace{1cm} (2)

where $w_{kl}$ refers to the weight between the $k^{th}$ output neuron and the $i^{th}$ input neuron, $\eta(t)$ is a learning rate which decreases slowly as a function of iteration number ($t$), the function $\Lambda(|y - y^*|)$ is called the “window function” and has the value of 1.0 when the $k^{th}$ output neuron is the neuron producing the maximum activation result $y^*$ (Duda et al., 2001). Some parameters may affect the performance of SOM neural networks, including the SOM map size, the initial neighborhood radius, the minimum learning rate, and the maximum learning rate (Li and Eastman, 2006).

4.2.1.3 ART neural networks

As recurrent networks, ART networks are powerful to recognize unexpected patterns and remember them for future use through feedback connections (Duda et al., 2001). The simplest ART network structure includes an input layer, a hidden layer and an output layer (Figure 4.3). Inputs of hidden neurons come from three sources: data from the input layer, feedback signals from the output layer through top-down weights $\hat{w}$, and a time-varying bias signal from the gain-
Figure 4.2 A Kohonen’s SOM neutral network with a two-dimension output layer. Input neurons are fully connected with output neurons through weighted links. Spatial relations among output neurons play an important role in neural networks training and pattern recognition.
Figure 4.3  Structure of ARTMAP neural networks. Different from feed-forward neural networks, feed-back signals are sent to hidden neurons, as well as the input data.
control unit. The gain-control system is a mechanism to keep a constant level of activity in the hidden layer, i.e., to hold $\|Y\|$ constant, where $Y$ refers to input vectors in the hidden layer (Duda et al., 2001). Outputs of hidden neurons flow to the output layer through down-top weights $w$. The down-top weights ($w$) represent the long-term memory of a cluster center while the top-down weights ($\tilde{w}$) provide an expectation for what comes from the bottom layer (Duda et al., 2001). The match between the input data and the learned clusters needs to be assessed. If the match is poor, then the cluster unit (output neuron) is suppressed by a reset signal (Figure 4.3). The criterion of an acceptable match is specified by a user-defined vigilance parameter $\rho$, which regulates the width of a cluster (Muchoney and Williamson, 2001).

Fuzzy ARTMAP neural networks are a supervised ART classification method. This method includes two ART modules that create stable pattern recognition based upon arbitrary sequences of training samples (Carpenter et al., 1997). Three parameters may influence the performance of fuzzy ARTMAP neural networks: the vigilance parameter $\rho$ ranging from 0 to 1, the learning rate parameter $\beta$ varying between 0 to 1, and the choice parameter $\alpha$ (Carpenter et al., 1997). The choice function ($\alpha$) selects the non-reset class with the highest activation that is determined by the size of the weight vector (Muchoney and Williamson, 2001). In general, a small choice parameter may lead to broad categories in the feature space while a large choice parameter corresponds to tight categories (Muchoney and Williamson, 2001). Compared to other neural networks, ART neural networks have been considered to be less sensitive to training parameters and over fitting (Lippitt et al., 2008).

4.2.1.4 Probabilistic neural networks

Probabilistic neural networks (PNN) calculate the probability of a pixel belonging to output classes based on the weighted sum of the input vector (Ashish et al., 2004; Zhang et al., 2009). PNNs are composed of an input layer (e.g. image bands), a hidden layer, and an output layer (desired classes) (Figure 4.4). Beginning with none of hidden layer neurons, PNNs create a hidden layer neuron whenever a new training sample is added. Weights between the created hidden neuron and input neurons are the transpose of the normalized input vector. Every hidden neuron is connected to only one output neuron that corresponds to the actual class of the training sample that creates this hidden neuron. The total number of hidden neurons is, therefore, equal to
Figure 4.4  Structure of probabilistic neural networks (PNNs)
the number of training samples. A nonlinear activation function is needed to transform the weighted sum of input data and produce outputs of hidden neurons. A commonly used function is the Gaussian activation function that can be written by:

\[ f(X) = e^{(W_kx - 1)/\sigma^2} \]  

(3)

where \( x \) refers to input vector; \( W_k \) is the weight vector of sample \( k \); and \( \sigma \) is a parameter (spread constant) that determines the size of the contribution of the weighted sum of input data to the output of activation function (Duda et al., 2001).

In brief, probabilistic neural networks are fully connected between the input and hidden layers, but sparsely connected between the hidden and output layers. In classification, each hidden neuron contributes to its associated class a value equal to the probability the test pixel was generated by a Gaussian centered on the associated training sample (Duda et al., 2001). The sum of these values reflects the probability of the test pixel belonging to the specific class. The training of PNN networks is very fast because the learning rule is very simple and requires only a single pass through training data. However, a large amount of memory is required because of the size of the hidden layer. In addition, it is very easy to add a new training sample into a trained probabilistic neural network. The setting of a user-defined parameter in activation function, namely, spread constant (\( \sigma \)), may affect the performance of PNN networks.

4.2.2 Network topology

The topology of a neural network is critical for neural computing to solve problems with reasonable training time and satisfactory performance. As the most commonly used neural networks, the topology of the MLP neural networks is determined by the number of hidden layers, neurons and connections, and the type of activation function.

4.2.2.1 Number of hidden layers, neurons and connections

The complexity of neural networks is largely defined by the total number of input features, the number of output classes, and the number of hidden layers and neurons (Duda, et al., 2001). While the first two factors are less flexible for a particular classification problem, we can adjust
the last two to find an appropriate size of neural networks. A trade-off is needed to balance the processing purpose of hidden layers and the training time needed. Large neural networks with more hidden layers may be more effective to represent non-linear complex relationships, but usually demand more training samples and longer training time. And larger neural networks are more likely to get caught in undesirable local minima or overly fit training data. On the other hand, compact neural networks may overly simplify the phenomena and lead to unsatisfactory results although they may be easier to train. Several prior studies investigated the issue of the optimal number of hidden layers, but their recommendations are not consistent. For example, Shupe and Marsh (2004) recommended single-hidden-layer networks while Civco (1993) preferred two-hidden-layer networks. Kanellopoulos and Wilkinson (1997) suggested that single-hidden-layer networks are suitable for most classification problems but two-hidden-layer networks may be more appropriate for those applications with more than 20 output classes. Some other researchers concluded that the number of hidden layers may not have a significant influence upon classification accuracies (e.g., Foody and Arora, 1997; Kavzoglu and Mather, 2003).

The number of neurons for the input, hidden, and output layers determines the number of weighted links, particularly for a fully connected network. Since the weight for each link is optimized through neural training, more links tend to increase the training time. Thus, every effort should be made to minimize unnecessary neurons in order to improve the efficiency in network training. Various feature extraction techniques, such as principal component analysis and discriminant analysis, have been used to reduce the data dimensionality and hence the number of neurons in the input layer (e.g., Benediktsson and Sveinsson, 1997; Liu and Lathrop, 2002). While the number of output neurons can be defined according to the research objectives in a specific application, a challenging issue is to choose the number of neurons in hidden layers. If there are too few neurons in hidden layers, the network may be unable to approximate very complex functions because of insufficient degrees of freedom. On the other hand, if there are too many neurons, the network tends to have a large number of degrees of freedom which may lead to overtraining and hence poor performance in generalization (Rojas, 1996). Thus, it is crucial to find the ‘optimum’ number of neurons in hidden layers that adequately capture the relationship
in training data. This optimization can be achieved by using trial and error or several systematic approaches such as pruning and constructive algorithms (Reed, 1993).

4.2.2.2 Activation function

Activation function is an algorithm that transforms the weighted sum of inputs and produces outputs of a neuron. Duda, et al. (2001) suggested a non-linear activation function should be used to deal with non-linear relationships; otherwise, neural networks would provide no computational power above linear classifiers. The commonly-used non-linear activation functions include the logistic sigmoid (log-sig) function (Equation 4) and the tangent sigmoid (tan-sig) function (Equation 5).

\[
f_1(x) = \frac{1}{1 + e^{-(1-a)x}} \tag{4}
\]

\[
f_2(x) = \frac{e^{(1-a)x} - e^{-(1-a)x}}{e^{(1-a)x} + e^{-(1-a)x}} \tag{5}
\]

where \(f_1(x)\) refers to log-sig activation function; \(f_2(x)\) is tan-sig activation function; \(x\) refers to the input vector; and \(a\) is an user-defined training threshold.

Several researchers examined different activation functions and concluded that the tan-sig function performed better (e.g., Ozkan and Erbek, 2003; Shupe and Marsh, 2004). However, little research has been conducted to investigate how the training threshold could affect the performance of these activation functions. Based on Equations (4)-(5), the training threshold determines the size of the contribution of input data to the output of a neuron. That is, it defines the slope of activation functions in their midrange (Figure 4.5). Therefore, the same type of activation function with various training thresholds may perform differently in image classification.

4.2.3 Neural networks training

Training is a learning process by which connection weights are adjusted until the network is deemed to be optimal. This involves the use of training samples, an error measure, and a learning algorithm. Training samples are presented to the network with input and output data over many
Figure 4.5  Curves of activation functions as related to different threshold (T) values, given that c is any constant between 0 and 1: (A) log-sig function and (B) tan-sig function.
iterations; they should not only be large in size but also be representative to ensure sufficient generalization ability. There are several different error measures, such as mean squared error (MSE), mean squared relative error (MSRE), root mean squared error (RMSE), and coefficient of efficiency (CE). The MSE has been most commonly used (Dawson and Wilby, 2001). There is a rich pool of training algorithms that vary by their origins, local or global perspectives, or learning modes. They are rooted in various techniques, such as optimum filtering, neurobiology, or statistical mechanics. Local learning algorithms adjust weights by using localized input signals and localized derivatives of the error function, while global algorithms consider all input signals. Based on the learning mode, training algorithms can be grouped into either a supervised, unsupervised, or hybrid paradigm. Reviews on these algorithms are given elsewhere (e.g., Bishop, 1995; Jain et al., 1996; Haykin, 1999; Principe et al., 2000). Here, I focused on three groups of supervised training methods that have been commonly used to train MLP neural networks: back-propagation, conjugate gradient, and Quasi-Newton methods; they differ in the ways by which the direction and magnitude of weight adjustments are calculated (Table 4.2).

4.2.3.1 Back-propagation methods

Due to its transparency and effectiveness, the back-propagation method has been widely used in neural network training (Duda et al., 2001; Kisi, 2007). It uses heuristic techniques adjusting the weights of neural networks along the negative of the gradient of the performance function. On each iteration, the derivatives of the classification error as functions of the weights are computed to determine the direction and magnitude of the weight update. As a result, the weights of neural networks are gradually optimized. Several commonly used back-propagation training algorithms include the steepest gradient descent (SGD), the gradient descent with momentum (GDM), and the resilient propagation (RP) algorithms.

Steepest gradient descent (SGD)

The SGD algorithm is the simplest back-propagation algorithm, which determines the direction and magnitude of the weights by the derivatives of the classification error with respect to any weight. Given a MLP neural network with one hidden layer, the weight update is computed
Table 4.2  List of some training algorithms that are particularly tied with a feed-forward network.

<table>
<thead>
<tr>
<th>Training Methods</th>
<th>Training Algorithms</th>
<th>Acronym</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-propagation</td>
<td>Steepest gradient descent</td>
<td>SGD</td>
<td>It determines the direction and magnitude of the weights by the derivatives of the classification error.</td>
</tr>
<tr>
<td></td>
<td>Gradient descent with momentum</td>
<td>GDM</td>
<td>This is an advanced gradient descent approach by adding a momentum item in the SGD weight adjustments formula.</td>
</tr>
<tr>
<td></td>
<td>Resilient propagation</td>
<td>RP</td>
<td>It attempts to speed up the training using an adoptive variable to define the magnitude of the weight update while the direction of the weight update is defined by the sign of the derivatives of the classification error.</td>
</tr>
<tr>
<td>Conjugate Gradient</td>
<td>Fletcher-Reeves</td>
<td>CGF</td>
<td>It calculates the conjugate of the previous search direction using the Fletcher-Reeves algorithm and then employs a line search to determine the optimal size of the weight update.</td>
</tr>
<tr>
<td></td>
<td>Polak-Ribiere</td>
<td>CGP</td>
<td>It is similar to the CGF method, differing only in the conjugate direction computation. It update the conjugate direction using the Polak-Ribiere algorithm.</td>
</tr>
<tr>
<td></td>
<td>Powell-Beale</td>
<td>CGB</td>
<td>Using the same learning model used by the CGF algorithm to compute the conjugate direction, it, however, resets the search direction to the negative of the gradient.</td>
</tr>
<tr>
<td></td>
<td>Scaled conjugate gradient</td>
<td>SCG</td>
<td>It is a variation of the conjugate gradient method with a scaled step size. It combines the model-trust region approach with the conjugate gradient approach to scale the step size.</td>
</tr>
<tr>
<td>Quasi-Newton</td>
<td>Broyden, Fletcher, Goldfarb, and Shanno</td>
<td>BFG</td>
<td>This algorithm approximates the Hessian matrix by a function of the gradient to reduce the computational and storage requirements.</td>
</tr>
<tr>
<td></td>
<td>Levenberg-Marquardt</td>
<td>LM</td>
<td>This method is considered as a combination of gradient descent and the Gauss-Newton method. It locates the minimum of a function that is expressed as the sum of squares of nonlinear functions.</td>
</tr>
</tbody>
</table>
through the chain-rule of partial derivatives. Specifically, the adjustment of the weight ($\Delta \omega$) can be calculated using:

$$\Delta \omega = -\eta \frac{\partial E}{\partial \omega}$$  \hspace{1cm} (6)

where $E$ is the classification error on this iteration, $\eta$ refers to the learning rate, and $\omega$ is the weights.

The SGD algorithm incorporates a user-defined learning rate. It is difficult to choose an appropriate learning rate. If the learning rate is set large, the algorithm may end up with oscillations around the minimum error; otherwise, the training may be inefficient. In addition, a fixed learning rate causes small changes in the weights even though the weights are far from their optimal values. Consequently, this algorithm usually takes more iterations to converge.

Gradient descent with momentum (GDM)

This algorithm is an advanced gradient descent approach adding a momentum item in the weight adjustment formula of the SGD algorithm to improve training efficiency. Momentum is a variable used to reduce the training time and avoid oscillations around the minimum. While the SGD algorithm adjusts the weights along the steepest gradient descent of the performance function, the GDM algorithm concerns both the current gradient descent and the recent changes of the weights. Therefore, the GDM algorithm can use a larger learning rate to speed up the training process with lower risk of oscillations. The weight adjustment equation on iteration $r$ by the GDM algorithm is given by:

$$\Delta \omega^r = -\eta \frac{\partial E}{\partial \omega} + \alpha \Delta \omega^{r-1}$$  \hspace{1cm} (7)

where $\alpha$ is a user-defined momentum factor ranging from 0 to 1. The value of momentum ($\alpha$) determines the size of the contribution of the most recent weights update on the current weights adjustment. A higher momentum value leads to a larger training step and then a faster training.

Resilient propagation algorithm (RP)
The SGD and GDM algorithms consider both the magnitude and the sign of the gradient of the performance function. A very small magnitude can cause small changes in the weights even though the weights are far from the optimal values. Consequently, these two algorithms usually take a long time to converge. The resilient propagation (RP) algorithm proposed by Riedmiller and Braun (1993) attempts to speed up the training process by eliminating the negative impacts of the magnitudes of the partial derivatives. It uses the sign of the derivatives of the classification error to determine the direction of the weight update. The magnitude of the weight update is defined by an individual variable which increases when the derivatives for two successive iterations have the same sign, decreases otherwise. The RP algorithm is in general much faster than other back-propagation training algorithms.

4.2.3.2 Conjugate Gradient Methods

The basic back-propagation method adjusts the weights along the steepest descent direction. However, this can be quite time-consuming due to the fixed learning rate. Different from the back-propagation method, the conjugate gradient method utilizes standard numerical optimization techniques, thus allowing very fast convergence. Rather than using the magnitude of the gradient descent and a user-defined learning rate, this method employs a series of line searches along conjugate directions to determine the optimal step size of the weight update. Specifically, the conjugate gradient training comprises two steps. The first step is to search the local minimum in error along a certain search direction; the weights will be adjusted to the local minimum point. The second step is to compute the conjugate of the previous search direction as the new search direction; the global minimum error is approached through iterations.

The conjugate direction $\Delta \omega^r$ on iteration $r$ can be written by combining the negative of gradient descent and the previous descent direction (Duda et al., 2001):

$$\Delta \omega^r = -\eta \frac{\partial E}{\partial \omega} + \beta_r \Delta \omega^{r-1}$$

where the $\beta_r$ can be computed by using either the Fletcher-Reeves algorithm or the Polak-Ribiere algorithm. This group of training method includes several commonly used algorithms,
such as Fletcher-Reeves (CGF), Polak-Ribiere (CGP), Powell-Beale (CGB), and scaled conjugate gradient (SCG) algorithms (see Table 4.2).

\textit{Fletcher-Reeves algorithm (CGF)}

This algorithm was proposed by Fletcher and Reeves (1964). It computes $\beta_r$ in Equation (8) as the ratio of the norm squared of the current gradient $\triangledown E(\omega^r)$ to the norm squared of the previous gradient $\triangledown E(\omega^{r-1})$:

$$
\beta_r = \frac{\triangledown E(\omega^r)^T \triangledown E(\omega^r)}{\triangledown E(\omega^{r-1})^T \triangledown E(\omega^{r-1})}
$$

(9)

where the superscript $T$ refers to the transpose of a matrix.

\textit{Polak-Ribiere algorithm (CGP)}

The CGP algorithm is similar to the CGF algorithm, but more robust in non-quadratic error functions than the CGF algorithm. It computes $\beta_r$ as the inner product of the previous change in the gradient with the current gradient divided by the norm squared of the previous gradient.

$$
\beta_r = \frac{\triangledown E(\omega^r)^T (\triangledown E(\omega^r) - \triangledown E(\omega^{r-1}))}{\triangledown E(\omega^{r-1})^T \triangledown E(\omega^{r-1})}
$$

(10)

\textit{Powell-Beale algorithm (CGB)}

Generally, if the error function is quadratic, the convergence of conjugate gradient descent is guaranteed when the number of iterations equals the number of network parameters (\textit{i.e.}, the total number of weights ($n$)). In practice, the error function may be non-quadratic, and we have to reset the search direction to the negative of the gradient. The CGB algorithm is using the same learning model to compute the conjugate direction with that used by the CGF algorithm (see Equations (8)-(9)). The CGB algorithm, however, resets the search direction using a method proposed by Powell (1977) while the CGF algorithm restarts the search every $n$ iterations. The CGB algorithm requires a slightly larger memory space than other conjugate gradient algorithms because of the computation of the restarting procedure.

\textit{Scaled conjugate gradient (SCG)}
Proposed by Moller (1993), the SCG algorithm is a variation of the conjugate gradient method with a scaled step size. The weights are still adjusted along the conjugate directions. To avoid the time-consuming line search on every iteration, however, the SCG algorithm combines the model-trust region approach with the conjugate gradient approach to scale the step size. The technical detail about how to scale the step size can be found in Moller’s (1993) article. This algorithm may require more iterations to converge than other conjugate gradient algorithms, but it usually involves less computation complexity and requires less computer memory for each iteration since no line search is needed.

4.2.3.3 Quasi-Newton methods

The Quasi-Newton method is an improved optimization method developed from Newton’s method that uses standard numerical optimization techniques to train neural networks (Demuth et al., 2008). The Newton's method uses standard numerical optimization techniques to train neural networks. The basic formula for the weight update through Newton's method is:

\[ \Delta \omega^r = -(H^r)^{-1} \nabla E(\omega^r) \quad (11) \]

where \((H^r)^{-1}\) is the inversion of the Hessian matrix on iteration \(r\). Although Newton's method can usually converge faster than conjugate gradient method, there are two major drawbacks. Firstly, it requires computing, storing, and inverting the Hessian matrix, which rapidly increases the computation complexity and requires a large memory space. On the other hand, Newton's method may not converge in non-quadratic error surfaces (Duda et al., 2001). The Quasi-Newton method updates an approximate Hessian matrix on every iteration, which is a function of the gradient, instead of the second-order derivatives matrix used by Newton's method. This update substantially reduces the computational complexity. When comparing to other methods, the Quasi-Newton method is more suitable for small networks with limited number of weights. The most successful Quasi-Newton method is the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm.

BFGS algorithm
This algorithm approximates the Hessian matrix by a function of the gradient to reduce the computational and storage requirements. The update of the approximation to the Hessian matrix on iteration $r$ can be written by:

$$
\Delta B^r = \frac{(\nabla E(\omega^r) - \nabla E(\omega^{r-1}))(\nabla E(\omega^r) - \nabla E(\omega^{r-1}))^T}{(\nabla E(\omega^r) - \nabla E(\omega^{r-1}))^T \Delta \omega^{r-1}} - \frac{B^{r-1} \Delta \omega^{r-1} (B^{r-1} \Delta \omega^{r-1})^T}{(\Delta \omega^{r-1})^T B^{r-1} \Delta \omega^{r-1}}
$$

(12)

where $\Delta B^r$ refers to the update of the approximation to the Hessian matrix on iteration $r$; and $B^{r-1}$ is the approximation to the Hessian matrix on iteration $r-1$.

When compared with conjugate gradient method, BFGS algorithm requires more complex computations and larger memory usage on every iteration. But this algorithm generally converges in fewer iterations since no line search is needed (Demuth et al., 2008).

**Levenberg-Marquardt algorithm (LM)**

The Levenberg-Marquardt (LM) algorithm is considered as a Quasi-Newton method since it also seeks for second-order training speed by using an approximate Hessian matrix. This algorithm approximates the Hessian matrix using the formulas:

$$
H \approx J^t J, \text{ and}
$$

$$
\text{gradient} = J^t e
$$

(13)

(14)

where $J$ is the Jacobian matrix that contains the derivatives of the performance function with respect to the weights; and $e$ is a vector of the network errors.

The weight update by this algorithm can be written by:

$$
\Delta \omega^r = -(J^t J + \mu I)^{-1} J^t e
$$

(15)

where $\mu$ is a scalar which decreases after each successful step (reduction in the performance function) and increases only when a tentative step would increase the performance function (Demuth et al., 2008). Since the memory space required for computing and storing Jacobian
matrix is proportional to the square of the number of the weights, this algorithm is restricted to small networks.

4.3 Internal parameters and classification accuracy

4.3.1 Experimental design

In this focused study, I assessed a set of topological and training parameters affecting image classification accuracy by the MLP neural networks. Three of the four parameters controlling the MLP network topology were considered, including number of hidden layers, type of activation function, and training threshold; the number of neurons was excluded since prior studies have found that image classification accuracies were less sensitive to this factor (e.g., Gong et al., 1996; Foody and Arora, 1997; Paola and Schowengerdt, 1997; Shupe and Marsh, 2004). The gradient descent with momentum (GDM) algorithm was used to train the MLP networks, and three related training parameters were considered, namely, learning rate, momentum, and number of iterations. These six internal parameters considered are summarized in Table 4.3.

The entire experiment comprised several major components. Firstly, I carefully constructed and trained a set of MLP neural network models with different topologies and training parameters. Then, I used these models to classify a satellite image into several major land cover categories, and I evaluated the accuracy of each classified map. Based on the classification accuracies, I further analyzed the sensitivity of these algorithm factors. Secondly, I compared the classification accuracies achieved by using the best neural network model and the Gaussian Maximum Likelihood (GML) classifier. Finally, I summarized our major findings and recommended several practical guidelines when parameterizing the MLP neural networks for image classification.

4.3.2 Remotely sensed data and land cover classification scheme

The remote sensor data used was a Landsat Enhanced Thematic Mapper Plus (ETM+) image dated on September 9th, 1999, which was geo-referenced to the UTM map projection. The image covers the northern Atlanta metropolitan area, Georgia, USA. The landscape in this area is characterized by a mosaic of urban use, agricultural use, and natural lands, making it an excellent
Table 4.3 List of the internal network parameters tested.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hidden layers</td>
<td>HL</td>
<td>A key factor controlling the topology of neural networks</td>
<td>$0 \leq HL \leq 4$</td>
</tr>
<tr>
<td>Activation function</td>
<td>AF</td>
<td>A linear or non-linear function for processing input data of neurons</td>
<td>Log-sigmoid or</td>
</tr>
<tr>
<td>Training threshold</td>
<td>TT</td>
<td>A user-defined threshold determining the contribution of the input data to</td>
<td>$0 \leq TT &lt; 1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the outcome</td>
<td></td>
</tr>
<tr>
<td>Learning rate</td>
<td>LR</td>
<td>A user-defined parameter defining the step size of the weight update</td>
<td>$0.001 \leq LR \leq 0.3$</td>
</tr>
<tr>
<td>Momentum</td>
<td>MO</td>
<td>A user-defined parameter controlling the influence of previous weight update</td>
<td>$0 \leq MO &lt; 1$</td>
</tr>
<tr>
<td>Number of iterations</td>
<td>IT</td>
<td>A parameter specifying how many times the training algorithm may iterate</td>
<td>$400 \leq IT \leq 2800$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>toward the targeted training goal</td>
<td></td>
</tr>
</tbody>
</table>

* Gradient descent with momentum (GDM) algorithm was used for neural training in this focused study.
site to test the effectiveness of different neural network configurations in image classification. In addition to the ETM+ image, I collected ancillary data through GPS-guided field observations and the use of high-resolution images available from the Google Earth.

Based on the remote sensor image and the ancillary data, I designed a mixed Anderson Level I/II land use/cover classification scheme (Anderson et al., 1976) with the following six major classes:

- High-density urban use: mostly large commercial and industrial buildings, large transportation facilities, and high-density residential areas in the city cores;
- Low-density urban use: mostly single/multiple family houses, apartment complexes, yards, local roads, and small open spaces;
- Exposed land: mainly non-impervious areas with sparse vegetation, such as clear-cuts, quarries, barren rock or sand along river/stream beaches;
- Cropland/grassland: crop fields and pasture as well as cultured grasses (such as golf courses, lawns, city parks);
- Forest: deciduous, coniferous, and mixed forest land; and
- Water: streams, rivers, lakes, and reservoirs.

4.3.3 Network configuration and training

I carefully configured 53 MLP neural network models with different internal parameters combinations. For each neural network model, the input neurons comprised seven ETM+ image bands (excluding the thermal band due to the coarse spatial resolution) and the output neurons were six major land use/cover classes. The general rule is that for the six internal parameters, only one parameter is allowed to alter at one time while holding others unchanged. In this way, the performance of neural networks with respect to a specific internal parameter can be assessed. Specifically, to investigate the impact of hidden layer number, five neural network models were constructed with the number of hidden layers changing from 0 to 4 (Table 4.4, No.1-5). Then, twenty neural network models were constructed to address the issues of activation function and training threshold (Table 4.4, No. 6-25). Both log-sig function and tan-sig function were
Table 4.4  List of the 53 neural network models configured with various internal parameter settings and classification accuracies obtained. Note that Mode 5 failed to converge during the training phase.

<table>
<thead>
<tr>
<th>No.</th>
<th>HL</th>
<th>AF</th>
<th>TT</th>
<th>LR</th>
<th>MO</th>
<th>IT</th>
<th>Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>79.33</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>80.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>77.67</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>66.67</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>82.00</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>log-sig</td>
<td>0.1</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>79.67</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>log-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>80.00</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>log-sig</td>
<td>0.3</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>80.00</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>log-sig</td>
<td>0.4</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>80.00</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>log-sig</td>
<td>0.5</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>79.67</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>log-sig</td>
<td>0.6</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>79.00</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>log-sig</td>
<td>0.7</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>78.33</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>log-sig</td>
<td>0.8</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>79.00</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>log-sig</td>
<td>0.9</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>72.00</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>tan-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>34.67</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>tan-sig</td>
<td>0.1</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>34.00</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>tan-sig</td>
<td>0.2</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>26.67</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>tan-sig</td>
<td>0.3</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>44.67</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>tan-sig</td>
<td>0.4</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>44.67</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>tan-sig</td>
<td>0.5</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>44.00</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>tan-sig</td>
<td>0.6</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>40.67</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>tan-sig</td>
<td>0.7</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>38.67</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>tan-sig</td>
<td>0.8</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>40.33</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>tan-sig</td>
<td>0.9</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>35.00</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.3</td>
<td>0.6</td>
<td>1000</td>
<td>53.67</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.25</td>
<td>0.6</td>
<td>1000</td>
<td>60.00</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.2</td>
<td>0.6</td>
<td>1000</td>
<td>75.00</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.15</td>
<td>0.6</td>
<td>1000</td>
<td>64.67</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.10</td>
<td>0.6</td>
<td>1000</td>
<td>72.33</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.05</td>
<td>0.6</td>
<td>1000</td>
<td>75.33</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>82.00</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.005</td>
<td>0.6</td>
<td>1000</td>
<td>81.33</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.001</td>
<td>0.6</td>
<td>1000</td>
<td>75.00</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>1000</td>
<td>81.33</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.1</td>
<td>1000</td>
<td>81.33</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.2</td>
<td>1000</td>
<td>81.33</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.3</td>
<td>1000</td>
<td>82.67</td>
</tr>
<tr>
<td>39</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.4</td>
<td>1000</td>
<td>82.00</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.5</td>
<td>1000</td>
<td>81.33</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.6</td>
<td>1000</td>
<td>82.00</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.7</td>
<td>1000</td>
<td>83.33</td>
</tr>
<tr>
<td>43</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>1000</td>
<td>84.00</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.9</td>
<td>1000</td>
<td>82.33</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>400</td>
<td>83.00</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>700</td>
<td>83.33</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>1000</td>
<td>84.00</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>1300</td>
<td>84.67</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>1600</td>
<td>84.00</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>1900</td>
<td>83.33</td>
</tr>
<tr>
<td>51</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>2200</td>
<td>81.67</td>
</tr>
<tr>
<td>52</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>2500</td>
<td>81.00</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>log-sig</td>
<td>0</td>
<td>0.01</td>
<td>0.8</td>
<td>2800</td>
<td>80.33</td>
</tr>
</tbody>
</table>

* Full names of the parameters are given in Table 4.3.
considered, and each was combined with a set of training threshold values ranging from 0 to 0.9. Finally, twenty-eight neural network models were configured to assess the three training parameters, and the range and step of each training parameter are listed in Table 4.4.

Each of the 53 neural network models was trained with an identical training sample set that contains 250 pixels for each land cover class. The training performance was measured by the root mean square error (RMSE). The training goal was set to 0.1 in RMSE. The training process stopped when either the maximum number of iterations or the training goal was reached. Most of the neural network models successfully converged except the model with four hidden layers (Table 4.4, No. 5).

4.3.4 Image classification and accuracy assessment

Each trained neural network model was used to classify the ETM+ image into the six land cover classes and hence a total of 52 land cover maps were produced with the exception of Model 5 that the training process failed to converge (Table 4.4, No. 5). The classification accuracy of each map was assessed by using the confusion matrix method that is based on the use of a reference dataset (Congalton, 1991). It computes the overall accuracy, user’s accuracies, producer’s accuracies, and the Kappa statistic through the comparison of the predicted values and the actual values of the reference samples. Kappa statistics are the indices of agreement for categorical data developed by Cohen, which were subsequently adopted by the remote sensing community as a useful measure of classification accuracy (Rossiter, 2004). When compared to overcall accuracy, user’s accuracies, and producer’s accuracies, Kappa coefficients consider not only the correctly classified reference data, but also the differences in the emergence chance of these land cover classes caused by the differences of their actual coverage. Therefore, with a small number of land cover classes, Kappa coefficients can provide a better assessment of classification accuracy by considering the emergence chance of different classes. In this research, to correctly perform the accuracy assessment, a reference dataset was collected by using the stratified random sampling scheme with approximately 50 samples for each land cover class. For each of the 52 land use/cover maps, a confusion matrix was created, and the overall classification accuracies were used to evaluate the performance of each neural network model for land cover classification.
For comparison purposes, the Gaussian Maximum Likelihood (GML) classifier was also trained with the identical training samples and then used to produce a land use/cover map from the ETM+ image. The classification accuracy was assessed with the same reference samples. I further compared the classification accuracies achieved by using the best neural network model (Table 4.4, No. 48) and the Gaussian Maximum Likelihood (GML) classifier.

4.3.5 Interpretation and analysis

4.3.5.1 Hidden layer number and classification accuracy

Based on Table 4.4 (No. 1-5) and Figure 4.6A, it is found that the performance of neural networks was quite sensitive to the hidden layer number, as indicated by the relatively high standard deviation. Among these models, the one with single hidden layer produced the best overall classification accuracy, followed by the models with zero hidden layer, two hidden layers, and three hidden layers. This finding concurs with those from Shupe and Marsh (2004) and Kanellopoulos and Wilkinson (1997). Theoretically, neural network models with more hidden layers can deal with more complex problems but require a large sample size to train. When input and output neurons are limited in number and training size is moderate or relatively small, neural network models equipped with more hidden layers can become less effective or even fail to converge in the training phase as they may end with local minima or overly fit the training data. Thus, the selection of an appropriate hidden layer number should consider the complexity of input and output neurons as well as the training sample size.

4.3.5.2 Activation function, training threshold, and classification accuracy

Table 4.4 (No. 6-25) and Figure 4.6B suggest that the activation function type can greatly affect the performance of neural network models. Clearly, the models equipped with the log-sig function substantially outperformed the ones with the tan-sig function, as indicated by the average overall accuracies. When incorporating training threshold in the comparison, I found that the models with the log-sig function were less sensitive to the training threshold values used when comparing to the ones equipped with the tan-sig function, as indicated by their standard deviations (2.65 versus 5.79). For the former group of models, the best classification accuracy was obtained when the training threshold value was set as 0. Although the variation of
Figure 4.6  Classification accuracies (Y-axis) as related to different internal settings (X-axis): (A) number of hidden layers, (B) activation function and training threshold, (C) learning rate, (D) momentum, and (E) number of iterations. For each figure, the average accuracy and the standard deviation (SD) are provided, and for Figures C-E, both raw data (solid lines) and linear trends (dash lines) with linear regression equations and R square values are shown.
classification accuracies by this group of models was relatively small, a decline trend emerged when the training threshold value was raised to 0.9. This suggests that a smaller training threshold value should be used for this group of models equipped with the log-sig function. For the models with the tan-sig function, the variation of their classification accuracies was larger, and relatively higher accuracies were obtained with moderate training threshold values (0.3-0.5).

4.3.5.3 Training parameters and classification accuracy

Table 4.4 (No. 26-53) and Figure 4.6 show the overall classification accuracies in relation to the three training parameters, namely, learning rate (Figure 4.6C), momentum (Figure 4.6D), and number of iterations (Figure 4.6E). Overall, neural network models were highly sensitive to the learning rate used, as indicated by the large standard deviation. The models with the learning rate ranging from 0.005 to 0.01 yielded higher classification accuracies. As the learning rate increased, the classification accuracy plunged by more than 20%, although another peak did occur when the learning rate increased to 0.2 (Figure 4.6C). Overall, increasing the momentum value helped boost the classification accuracy (Figure 4.6D), especially when this value was larger than 0.6; by adjusting the value of momentum, the classification accuracy increased from 81.33% to 84% (Table 4.4, No. 35-44). Nevertheless, the impact of momentum on classification accuracy was quite marginal, as indicated by the relatively small standard deviation. The number of iterations had a moderate impact upon the classification accuracy, as shown by the standard deviation. The overall classification accuracy increased as the number of iterations increased to 1300; after that the accuracy began to decline (Figure 4.6E).

4.3.5.4 MLP neural networks and Gaussian maximum likelihood classifier

Table 4.5 summarizes the classification accuracies by the best neural network model (Table 4.4, No. 48) and the Gaussian Maximum Likelihood (GML) classifier. Clearly, the neural network model showed a moderate improvement in the overall classification accuracy and the overall Kappa coefficient when comparing to the outcome by the GML classifier. When looking at the conditional Kappa coefficients for different land use/cover types, however, the neural network model performed much better when classifying the two spectrally complex urban classes, which further confirms the robustness of the neural network technique in dealing with non-linear,
Table 4.5  Comparison of classification accuracies by artificial neural networks (ANNs, Model 48 of Table 4.4) and Gaussian Maximum Likelihood (GML) classifier.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Conditional Kappa Coefficients</th>
<th>Overall Kappa Coefficient</th>
<th>Overall accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High-density urban</td>
<td>Low-density urban</td>
<td>Exposed land</td>
</tr>
<tr>
<td>ANNs</td>
<td>0.74</td>
<td>0.67</td>
<td>0.80</td>
</tr>
<tr>
<td>GML</td>
<td>0.61</td>
<td>0.52</td>
<td>0.92</td>
</tr>
</tbody>
</table>
complex phenomena. On the other hand, the GML classifier performed better in classifying several spectrally homogenous land use/cover classes, such as forest, cropland, or exposed land, confirming the applicability of this popular parametric classifier.

4.3.6 Summary

In this focused study, I investigated the sensitivity of neural networks to six topological and training parameters. I found that the performance of neural networks was highly sensitive to number of hidden layers, type of activation function, and training rate. And the three other parameters, *i.e.*, training threshold, momentum, and number of iterations, had a marginal impact upon the classification accuracy. A careful neural network configuration can lead to a moderate overall accuracy improvement and a substantial improvement for the two urban classes when comparing to the outcome by the Gaussian Maximum Likelihood (GML) classifier. These observations suggest the importance of internal parameter settings when using neural networks for image classification.

On the other hand, several practical guidelines emerged from this study, which can be useful when parameterizing the MLP neural network architecture for image classification. Specifically, a small number of hidden layers should be used when the training sample size is moderate or the number of input and output neurons is small. Using a large number of hidden layers will increase the computational complexity and lead to a suboptimal or unsatisfactory performance when the training sample size or the number of input and output neurons is not large enough. The log-sig function should be used as it can help yield much better classification accuracies and is relatively less sensitive to training threshold values. For better classification accuracies, a small learning rate, large momentum, and moderate number of iterations should be used.

4.4 Training algorithm performance

4.4.1 Experimental design

In this focused study, I evaluated the performance of several popular training algorithms in image classification by the MLP networks. They include steepest gradient descent (SDE), gradient descent with momentum (GDM), resilient propagation (RP), Fletcher-Reeves (CGF),
Polak-Ribiere (CGP), Powell-Beale (CGB), scaled conjugate gradient (SCG), BFGS (Broyden, Fletcher, Goldfarb, and Shanno) (BFG), and Levenberg-Marquardt (LM) algorithms (Table 4.2). I used each algorithm to train the MLP networks multiple times using identical training samples, and then applied each of the resultant network models to derive land cover information from the ETM+ image described earlier. The training algorithms were further evaluated according to their training efficiency, capability of convergence, classification accuracy, and stability of the classification accuracy.

4.4.2 Network training and image classification

I constructed a MLP network with seven input neurons, twenty neurons in a single hidden layer, and ten output neurons. The activation functions for the hidden and output neurons were hyperbolic tangent function and logistic sigmoid function, respectively. The input data were the seven ETM+ image bands excluding the thermal band because of its coarse spatial resolution, and the output layer consisted of ten land use/cover classes or sub-classes. While the land classification scheme used here was the same as the one described earlier, the high-density urban comprised three subclasses (i.e., open space, large roof building, and small roof building in the city core), cropland/grassland included two subclasses (i.e., well-vegetated grassland and less-vegetated land), and forest consisted of two subclasses (i.e., coniferous/mixed forest and deciduous forest). This is why the output layer comprised ten neurons. For each subclass/class, 250 pixels were collected as the training data, and the training performance was measured by the mean square error (MSE).

Several network training parameters were initiated before actually training. Specifically, the training goal was set at 0.03 in terms of MSE, the training time and iterations were infinite, and both minimum gradient and minimum step were defined as 1e-006. Note that the learning rate was set as 0.01 for the SGD algorithm or 0.02 for the GDM algorithm. The momentum factor for the GDM algorithm was defined as 0.6. The training process was stopped when the MSE error reached the training goal, indicating that the training successfully converged, or when the minimum gradient or the minimum step was met, showing that the training failed to converge. To minimize the impacts of the initial weights, I used each of the nine training algorithms to train the network ten times with the above training parameters settings. As a result, ninety network
models were created, which were further used to classify the ETM+ scene into ten land cover classes or subclasses which were finally merged into the six major land use/cover classes. In ninety land use/cover maps were produced.

4.4.3 Performance evaluation

The performance of each training algorithm was evaluated by using the four criteria, namely, training efficiency, convergence capability, classification accuracy, and stability of the classification accuracy.

4.4.3.1 Training efficiency

Different training algorithms vary in their computational intensity and the time to reach the training goal. An efficient training algorithm can considerably reduce the time cost for image classification by neural networks. Therefore, the training efficiency has been considered as a critical criterion for examining the usefulness of training algorithms (Skinner and Broughton, 1995). The number of iterations used for training is a good indicator of the efficiency of training algorithms (Kanellopoulos and Wilkinson, 1997; Kisi, 2007). Here, I used the average number of iterations during the ten experiments to quantify the training efficiency (Table 4.6).

From Table 4.6, it is clear that the training efficiency of different algorithms varied greatly. The LM algorithm was the most efficient. Several algorithms, such as RP, CGF, CGP, CGB, SCG, and BFG, showed a moderate training efficiency. Both the SGD and GDM algorithms were extremely poor in terms of the training efficiency.

4.4.3.2 Capability of convergence

The capability of convergence provides the information on how often a training algorithm can reach the training goal. Failure to converge usually leads to a poor classification performance. Therefore, the capability of convergence has been considered as an important criterion for measuring the performance of training algorithms (Skinner and Broughton, 1995; Kanellopoulos and Wilkinson, 1997). Here, the capability of convergence was defined as the rate of convergence, which is actually the ratio between the number of the converged experiments and the total number of the experiments by a specific algorithm (Table 4.6).
Table 4.6  Summary of the training algorithm performance.

<table>
<thead>
<tr>
<th>Training Algorithms</th>
<th>Mean Iterations</th>
<th>Convergence Rate (%)**</th>
<th>Classification Accuracy***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>Mean (%) Standard Deviation</td>
</tr>
<tr>
<td>SGD</td>
<td>65460</td>
<td></td>
<td>77.8 2.3</td>
</tr>
<tr>
<td>GDM</td>
<td>65432</td>
<td></td>
<td>76.1 3.3</td>
</tr>
<tr>
<td>RP</td>
<td>805</td>
<td></td>
<td>76.4 2.3</td>
</tr>
<tr>
<td>CGF</td>
<td>1452</td>
<td>30</td>
<td>61.7 16.6</td>
</tr>
<tr>
<td>CGP</td>
<td>865</td>
<td>100</td>
<td>75.1 4.4</td>
</tr>
<tr>
<td>CGB</td>
<td>627</td>
<td>50</td>
<td>64.5 17.8</td>
</tr>
<tr>
<td>SCG</td>
<td>1486</td>
<td>100</td>
<td>76.1 4.5</td>
</tr>
<tr>
<td>BFG</td>
<td>389</td>
<td>0</td>
<td>36.8 17.8</td>
</tr>
<tr>
<td>LM</td>
<td>15</td>
<td>100</td>
<td>77.7 2.7</td>
</tr>
</tbody>
</table>

* Full names of the training algorithms are given in Table 4.2.
** The convergence rate is defined as the ratio between the number of the converged experiments and the total number of the experiments for each training algorithm.
*** The results were based on ten experiments.
All the three back-propagation algorithms (SGD, GDM, and RP) successfully converged in every experiment. Two of the conjugate gradient algorithms, namely, CGP and SCG, and the Levenberg-Marquardt (LM) algorithm were also quite good in this regard. However, the other two conjugate gradient algorithms, namely, CGF and CGB, were quite poor in terms of their capability of convergence. The BFG algorithm failed to converge in all experiments, which may be due to the emergence of non-quadratic error surfaces.

4.4.3.3 Classification accuracy

The performance of a pattern classifier is usually assessed by estimating its classification accuracy. Training algorithms resulting in poor classification accuracies are less useful in practice. Here, I used the confusion matrix method described earlier to measure the classification accuracy. A test dataset with approximately 50 samples randomly selected for each land use/cover subclass/class was used for this purpose. The overall classification accuracies were compared to evaluate the performance of each training algorithm.

Based on Table 4.6, it is clear that the classification accuracy varies by training algorithms. If using the average classification accuracy as the reference, the SGD and LM algorithms clearly performed the best, followed by the GMD, RP, CGP, and SCG algorithms. The BFG, CGB, and CGF algorithms generated relatively lower average classification accuracies, which were mostly caused by their failure to converge in one or more experiments.

4.4.3.4 Stability of the classification accuracy

The stability of the classification accuracy was another criterion that I used to evaluate the performance of a training algorithm. The training methods which result in relatively stable classification accuracies are generally preferred. To evaluate the stability of the classification accuracy, the standard deviation of the classification accuracies from the ten experiments for each algorithm was computed. A smaller standard deviation suggests a more stable performance. Based on Table 4.6, it is found that the three back-propagation algorithms, namely, SGD, GDM, and RP, and the LM algorithm were quite stable, the CGP and SCG algorithms were less stable, and the CGF, CGB and BFG algorithms were least stable. The poor performance of the last group of training algorithms was clearly attributed to their poor capability of convergence.
4.4.3.5 Overall evaluation

When combining the four categories of comparison, it is found that the LM algorithm performed the best. It was the most efficient algorithm with a strong capability of convergence, providing the most accurate and stable land use/cover classification accuracies. However, this algorithm may become less efficient when a large memory space is needed to accommodate the high complexity of computation due to increasing network size. The RP algorithm performed almost identical to the LM algorithm except that the former used more iterations to converge. Since the RP algorithm does not require computing the matrix of the second-order derivatives of its performance function, the networks size increase will have a limited impact upon the performance. With a much lower computational cost and less memory usage, the RP algorithm can outperform the LM algorithm when training large networks with many parameters. The CGP and SCG algorithms also performed well although they showed moderate training efficiency and the resultant classification accuracies were not as stable as those from the LM and RP algorithms. The SGD and GDM algorithms were not competitive when compared to the LM, RP, CGP, and SCG algorithms due to their extremely low training efficiency. Finally, the CGF, CGB, and BFG algorithms were not recommended as they showed a poor capability of convergence which led to poor classification accuracies. The above observations suggest the importance of selecting an appropriate training algorithm when using artificial neural networks for land use/cover classification.

4.5 Neural network architectures assessment

4.5.1 Experimental design

Here I assessed the performance of several popular neural network architectures on land cover classification. They include MLP neural networks, Kohonen’s SOM neural networks, Fuzzy ARTMAP neural networks, and probabilistic neural networks (See Section 4.2.1 for their technical details). I firstly optimized parameters setting for each individual architecture by using trial-and-error experiments. Then, I applied each of these appropriately parameterized neural network architectures to derive land cover information from the ETM+ image described earlier.
The network architectures were further assessed according to their training efficiencies and classification accuracies.

### 4.5.2 Network configuration and image classification

The input data were seven bands of the ETM+ scene described earlier excluding the thermal band (see Section 4.3.3). Outputs were the six classes in the land use/cover classification scheme described in Section 4.3.2. The cross-validation method, which is a widely used technique for estimating the performance of image classifiers, was adopted to train neural networks and assess classification results. A total of 500 ground-truth pixels were collected for each class. They were randomly divided into two sets: one used for neural network training, and the other for classification accuracy assessment.

Based on the two focused studies I conducted (see Sections 4.3 and 4.4), I constructed a single-hidden-layer MLP network with the optimal parameter setting. Specifically, a logistic activation function with threshold 0 was adopted. Considering its easiness of implementation and robust performance in land cover classification, an advanced GDM algorithm was used for network training. Regarding training parameters, the training rate was set to 0.01; the momentum was 0.8; and the number of iterations was 1300.

As for the parameters setting of Kohonen’s SOM neural networks, Li and Eastman (2006) and Hu and Weng (2009) suggested that the minimum and maximum learning rates equal to 0.5 and 1 yield high classification accuracies, and that initial neighborhood radius should be large enough to cover the entire output layer. Li and Eastman (2006) also found that an extremely small gain term ranging from 0.0001 to 0.0005 led to satisfactory results. Nevertheless, few efforts have been made to investigate the optimal SOM map size (output neuron density) and tuning iterations. Here, I parameterized a set of SOM networks with various SOM map sizes and tuning iterations while other parameters were set as recommended by existing literature (Table 4.7). These networks were then applied to classifying the ETM+ scene into the six land cover classes (see Section 4.3.2). The optimal neuron density and tuning iterations were evaluated in terms of classification accuracies by these networks.
Table 4.7  Experimental results of land cover classification by SOM neural networks as related to the map size and fine tuning iterations.

<table>
<thead>
<tr>
<th>No.</th>
<th>Map Size (Output Neuron Density)</th>
<th>Fine Tuning Iterations</th>
<th>Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15*15</td>
<td>50</td>
<td>92.47</td>
</tr>
<tr>
<td>2</td>
<td>15*15</td>
<td>100</td>
<td>92.67</td>
</tr>
<tr>
<td>3</td>
<td>15*15</td>
<td>150</td>
<td>92.53</td>
</tr>
<tr>
<td>4</td>
<td>15*15</td>
<td>200</td>
<td>92.75</td>
</tr>
<tr>
<td>5</td>
<td>20*20</td>
<td>50</td>
<td>91.60</td>
</tr>
<tr>
<td>6</td>
<td>20*20</td>
<td>100</td>
<td>91.87</td>
</tr>
<tr>
<td>7</td>
<td>20*20</td>
<td>150</td>
<td>92.20</td>
</tr>
<tr>
<td>8</td>
<td>20*20</td>
<td>200</td>
<td>93.00</td>
</tr>
<tr>
<td>9</td>
<td>20*20</td>
<td>250</td>
<td>92.13</td>
</tr>
<tr>
<td>10</td>
<td>20*20</td>
<td>300</td>
<td>92.00</td>
</tr>
<tr>
<td>11</td>
<td>25*25</td>
<td>50</td>
<td>91.93</td>
</tr>
<tr>
<td>12</td>
<td>25*25</td>
<td>100</td>
<td>92.53</td>
</tr>
<tr>
<td>13</td>
<td>25*25</td>
<td>150</td>
<td>92.07</td>
</tr>
<tr>
<td>14</td>
<td>25*25</td>
<td>200</td>
<td>92.40</td>
</tr>
<tr>
<td>15</td>
<td>25*25</td>
<td>250</td>
<td>93.07</td>
</tr>
<tr>
<td>16</td>
<td>25*25</td>
<td>300</td>
<td>93.20</td>
</tr>
<tr>
<td>17</td>
<td>25*25</td>
<td>350</td>
<td>92.60</td>
</tr>
<tr>
<td>18</td>
<td>25*25</td>
<td>400</td>
<td>92.40</td>
</tr>
<tr>
<td>19</td>
<td>30*30</td>
<td>50</td>
<td>92.53</td>
</tr>
<tr>
<td>20</td>
<td>30*30</td>
<td>100</td>
<td>92.40</td>
</tr>
<tr>
<td>21</td>
<td>30*30</td>
<td>150</td>
<td>92.60</td>
</tr>
<tr>
<td>22</td>
<td>30*30</td>
<td>200</td>
<td>93.33</td>
</tr>
<tr>
<td>23</td>
<td>30*30</td>
<td>250</td>
<td>92.93</td>
</tr>
<tr>
<td>24</td>
<td>30*30</td>
<td>300</td>
<td>92.87</td>
</tr>
<tr>
<td>25</td>
<td>35*35</td>
<td>50</td>
<td>90.93</td>
</tr>
<tr>
<td>26</td>
<td>40*40</td>
<td>50</td>
<td>90.93</td>
</tr>
<tr>
<td>27</td>
<td>45*45</td>
<td>50</td>
<td>92.93</td>
</tr>
<tr>
<td>28</td>
<td>50*50</td>
<td>50</td>
<td>91.80</td>
</tr>
</tbody>
</table>

* Those experiments yielded relatively high classification accuracies.
To optimize ARTMAP neural networks, a set of ARTMAP neural networks were parameterized with the choice value ranging from 0 to 0.9 while keeping other parameters unchanged (see Table 4.8, No. 1-10). Another set of neural networks were configured with the vigilance value ranging from 0.1 to 1 while fixing other parameters (Table 4.8, No.10-23). Ten more neural networks were parameterized by changing the learning rate (Table 4.8, No. 23-32). These networks were used to classify the ETM+ image. The optimal parameter setting for ARTMAP neural networks was evaluated by comparing the classification accuracies by these models.

For PNNs parameterization, I constructed a series of probabilistic neural networks with the spread constant value ranging from 0.001 to 1 (Table 4.9). I firstly trained these PNNs models with an identical training set, and then classified the ETM+ image using these models and assessed their classification accuracies.

In total, four types of neural networks were configured with the optimal parameter settings identified through the above experiments and then applied for land cover classification from the ETM+ image. Their performances were further assessed in terms of the overall classification accuracy and the specific accuracy for the urban land classes.

4.5.3 Results and discussions

Based on Table 4.7, it is clear that 200-300 fine tuning iterations led to higher classification accuracies. With the map size of 15, 20 or 30, the iteration number 200 had the best performance. When the map size was 25, the optimal iteration number was 300. I also found that the increase of the map size slightly improved land cover classification, but the map size larger than 30*30 ended up with poor classification accuracies. The highest classification accuracy by the SOM neural networks (93.3%) was achieved by using the map size of 30 in combination with the iteration number of 200.

Table 4.8 shows that the performance of ARTMAP neural networks was very sensitive to the vigilance value but less sensitive to the choice value. Classification accuracies declined rapidly as the vigilance value decreased. In addition, all the experiments with the learning rate rather than 1 failed to classify the image. The highest classification accuracy by the Fuzzy ARTMAP
Table 4.8  Experimental results of land cover classification by Fuzzy ARTMAP neural networks as related to various settings for the three parameters: choice, vigilance, and learning rate.

<table>
<thead>
<tr>
<th>No.</th>
<th>Choice</th>
<th>Vigilance</th>
<th>Learning Rate</th>
<th>Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.98</td>
<td>1</td>
<td>91.80</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.98</td>
<td>1</td>
<td>91.80</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.98</td>
<td>1</td>
<td>91.80</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.98</td>
<td>1</td>
<td>91.80</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>0.98</td>
<td>1</td>
<td>91.80</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>0.98</td>
<td>1</td>
<td>91.73</td>
</tr>
<tr>
<td>7</td>
<td>0.6</td>
<td>0.98</td>
<td>1</td>
<td>91.93</td>
</tr>
<tr>
<td>8</td>
<td>0.7</td>
<td>0.98</td>
<td>1</td>
<td>92.13</td>
</tr>
<tr>
<td>9</td>
<td>0.8</td>
<td>0.98</td>
<td>1</td>
<td>92.13</td>
</tr>
<tr>
<td>10</td>
<td>0.9</td>
<td>0.98</td>
<td>1</td>
<td>91.93</td>
</tr>
<tr>
<td>11</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>16.73</td>
</tr>
<tr>
<td>12</td>
<td>0.7</td>
<td>0.96</td>
<td>1</td>
<td>89.40</td>
</tr>
<tr>
<td>13</td>
<td>0.7</td>
<td>0.94</td>
<td>1</td>
<td>84.67</td>
</tr>
<tr>
<td>14</td>
<td>0.7</td>
<td>0.92</td>
<td>1</td>
<td>81</td>
</tr>
<tr>
<td>15</td>
<td>0.7</td>
<td>0.9</td>
<td>1</td>
<td>69.40</td>
</tr>
<tr>
<td>16</td>
<td>0.7</td>
<td>0.8</td>
<td>1</td>
<td>45.60</td>
</tr>
<tr>
<td>17</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>39.47</td>
</tr>
<tr>
<td>18</td>
<td>0.7</td>
<td>0.6</td>
<td>1</td>
<td>26.93</td>
</tr>
<tr>
<td>19</td>
<td>0.7</td>
<td>0.5</td>
<td>1</td>
<td>22.53</td>
</tr>
<tr>
<td>20</td>
<td>0.7</td>
<td>0.4</td>
<td>1</td>
<td>19.67</td>
</tr>
<tr>
<td>21</td>
<td>0.7</td>
<td>0.3</td>
<td>1</td>
<td>19.07</td>
</tr>
<tr>
<td>22</td>
<td>0.7</td>
<td>0.2</td>
<td>1</td>
<td>17.53</td>
</tr>
<tr>
<td>23</td>
<td>0.7</td>
<td>0.1</td>
<td>1</td>
<td>16.67</td>
</tr>
<tr>
<td>24</td>
<td>0.7</td>
<td>0.98</td>
<td>0.98</td>
<td>N/A</td>
</tr>
<tr>
<td>25</td>
<td>0.7</td>
<td>0.98</td>
<td>0.96</td>
<td>N/A</td>
</tr>
<tr>
<td>26</td>
<td>0.7</td>
<td>0.98</td>
<td>0.94</td>
<td>N/A</td>
</tr>
<tr>
<td>27</td>
<td>0.7</td>
<td>0.98</td>
<td>0.92</td>
<td>N/A</td>
</tr>
<tr>
<td>28</td>
<td>0.7</td>
<td>0.98</td>
<td>0.9</td>
<td>N/A</td>
</tr>
<tr>
<td>29</td>
<td>0.7</td>
<td>0.98</td>
<td>0.8</td>
<td>N/A</td>
</tr>
<tr>
<td>30</td>
<td>0.7</td>
<td>0.98</td>
<td>0.7</td>
<td>N/A</td>
</tr>
<tr>
<td>31</td>
<td>0.7</td>
<td>0.98</td>
<td>0.6</td>
<td>N/A</td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>0.98</td>
<td>0.5</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 4.9  Experimental results of land cover classification by the probabilistic neural networks as related to varying spread constants.

<table>
<thead>
<tr>
<th>No.</th>
<th>Spread Constant</th>
<th>Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>88.67</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>88.67</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>88.67</td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>88.73</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>88.73</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>88.93</td>
</tr>
<tr>
<td>7</td>
<td>0.4</td>
<td>89.20</td>
</tr>
<tr>
<td>8</td>
<td>0.3</td>
<td>89.47</td>
</tr>
<tr>
<td>9</td>
<td>0.2</td>
<td>90.00</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>92.27</td>
</tr>
<tr>
<td>11</td>
<td>0.09</td>
<td>92.87</td>
</tr>
<tr>
<td>12</td>
<td>0.08</td>
<td>93.00</td>
</tr>
<tr>
<td>13</td>
<td>0.07</td>
<td>93.07</td>
</tr>
<tr>
<td>14</td>
<td>0.06</td>
<td>93.47</td>
</tr>
<tr>
<td>15</td>
<td>0.05</td>
<td>93.93</td>
</tr>
<tr>
<td>16</td>
<td>0.04</td>
<td>93.93</td>
</tr>
<tr>
<td>17</td>
<td>0.03</td>
<td>94.33</td>
</tr>
<tr>
<td>18</td>
<td>0.02</td>
<td>94.2</td>
</tr>
<tr>
<td>19</td>
<td>0.01</td>
<td>93.13</td>
</tr>
<tr>
<td>20</td>
<td>0.001</td>
<td>87.87</td>
</tr>
<tr>
<td>21</td>
<td>0.0001</td>
<td>17.73</td>
</tr>
</tbody>
</table>
neural networks (92.13%) was achieved when the choice value was 0.7 or 0.8, the vigilance value was 0.98, and the learning rate was 1.

From Table 4.9, it is found that the performance of PNNs varied as the spread constant value changed. In general, the classification accuracies by PNNs increased as the spread constant value decreased. PNNs had the best performance when the spread constant value was 0.03.

Clearly, MLP neural networks outperformed other three neural networks in terms of overall classification accuracies and the specific accuracies for the urban land use (Table 4.10 and Figure 4.7). PNNs and SOM networks had slightly lower overall classification accuracies, but much worse in terms of the classification accuracy for the urban land use. ARTMAP had the worst performance. The Kappa statistics by ARTMAP were only 0.7747 for high-density urban use and 0.8579 for low-density urban use while those by MLP were 0.8965 and 0.9077, respectively.

4.6 Conclusions

While the prospect of neural networks for image classification has been quite promising, the capability of neural networks tends to be oversold as an all-inclusive 'panacea' that is capable to outperform other classifiers regardless of network architecture, training algorithms, or data quality. Consequently, this field has been characterized by inconsistent research designs and immature operational practices. Nevertheless, recent studies suggested the need to adopt a systematic approach to pattern recognition by neural networks considering data acquisition and pre-processing, network configuration, training algorithms, and validation in a sequential mode (e.g., Mailer and Dandy, 2000; Principe et al., 2000; Mas and Flores, 2008; Yang, 2009a, b).

Here, I propose a systematic approach that can guide the use of neural networks for image classification from remote sensor data (Figure 4.8). It comprises several core technical components, beginning with data collection and acquisition that have been considered as a critical component in any remote sensing-based land use/cover mapping projects (Yang and Lo, 2002). At this stage, both remote sensor data and ancillary data should be collected to help prepare a land use/cover classification scheme, the training and validation data sets, and the primary data actually used in image classification. Every effort should be made to acquire
**Table 4.10** Comparisons of the optimal performances of the four types of neural networks for land cover classification.

<table>
<thead>
<tr>
<th>Neural Networks</th>
<th>Optimized Performance</th>
<th>Conditional Kappa Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall Classification Accuracy</td>
<td>High-density Urban</td>
</tr>
<tr>
<td>MLP</td>
<td>0.9620</td>
<td>0.8965</td>
</tr>
<tr>
<td>SOM</td>
<td>0.9333</td>
<td>0.8742</td>
</tr>
<tr>
<td>ARTMAP</td>
<td>0.9213</td>
<td>0.7747</td>
</tr>
<tr>
<td>PNNs</td>
<td>0.9433</td>
<td>0.8322</td>
</tr>
</tbody>
</table>
Figure 4.7  Illustration of the optimal performances of four types of neural networks in urban land cover classification.
Figure 4.8  A systematic approach to image classification by neural networks. While data acquisition and pre-processing are normally quite time consuming, neural network configuration is most challenging. Nevertheless, neural network training can be quite computationally intensive (Yang and Zhou, 2011).
sufficient data that represent the conditions that the neural network may encounter later. The information contents and data quality should be emphasized during the phase of data acquisition (Yang, 2009a).

Data pre-processing is necessary as it can have a significant effect on neural network performance (Foody et al., 1995; Foody and Arora, 1997; Mas and Flores, 2008). There are two major tasks during this phase. Firstly, it is necessary to determine which image bands should be actually included as the input data. If large image scenes or many image bands are being considered, a data dimensionality reduction technique like principal component analysis (e.g., Liu and Lathrop, 2002) should be used to extract salient features prior to the actual classification. This procedure can greatly reduce the computational burden. The second task is to identify the training, test, and validation datasets by using reference data in combination with some image interpretation procedures. The training dataset is for network training, the test dataset is used to assess the performance of the network at the training stage for cross-validation purposes, and the validation dataset is used to evaluate the performance of a network against independent data. Both the test and validation datasets should be much smaller in size when comparing to the training set but each dataset should be representative of the same population. If the available dataset is limited, the division of data may be difficult, and some other methods, such as bootstrapping (Kohavi, 1995) or the hold-out method (Masters, 1995), could be attempted to maximize utilization of the available data.

Prior to training, it is important to define an appropriate neural network architecture and training parameters. Begin with a multi-layer-perceptron neural network and a back-propagation learning algorithm as the benchmark to evaluate any other network types and learning methods. Specify an appropriate number of hidden layers and nodes unless a pruning algorithm or cascade correlation is used. Begin with one hidden layer as a starting point. Choose either logistic sigmoid or hyperbolic tangent function as the activation function. Also choose appropriate values for learning parameters. As demonstrated in the three focused studies, a number of trial-and-error experiments may need in order to optimize the network architecture. The initial weights should be randomly chosen. Use the training data set in the training, and the test set for cross-validation in order to determine when to terminate the training process. Neural networks training may adopt
a strategy to avoid overtraining by calculating the classification error of a test dataset on each iteration, and once the error goes up, stop training. In our focused studies, the conditions of stopping training were defined by the training goal and other parameters, such as minimum gradient size. Once the training is completed, saves the weights and architecture of the neural model, and applies the trained model to image classification to produce a land use/cover map. The classification performance is assessed by using the independent validation data through the error matrix method described earlier.
CHAPTER FIVE: URBAN GROWTH CHARACTERIZATION

5.1 Introduction

Measuring spatio-temporal dynamics of urban landscapes involves in several technical components, including image classification, urban change detection, urban growth pattern analysis, and urban growth process analysis (see Figure 3.2). These components need to be systematically organized in order to produce satisfactory results. In this chapter, I will firstly describe my approach for urban land change mapping, present my results, and discuss general trends of urban spatial growth in the Beijing metropolitan area. Next, I will discuss my landscape metrics analysis for urban growth characterization, examine the scale dependency of remote sensing-derived landscape metrics analysis, and discuss urban growth patterns and related urban growth processes. Finally, I will introduce a method for exploring the spatio-temporal dynamics of urban growth through quantitatively measuring the spatial variations of urban growth patterns using the moving windows analysis and GIS-based spatial analysis. These three approaches explored the features of urban spatial growth in the Beijing metropolitan area at the metropolitan, functional zone, and cell levels from different perspectives.

5.2 Urban land change mapping

Urban land change detection is useful for visualizing urban spatial expansion over time. Based upon produced urban land change maps, statistic summaries at the metropolitan and function zone levels further quantitatively measure the locations and rates of urban growth during the study period. Most importantly, urban land change detection provides base maps for further urban growth patterns and processes analysis. Thus, urban land change mapping from remote sensor data is an essential component of urban spatial growth characterization. Urban land change detection technique identified here included the following procedures: image preprocessing, classification scheme design, neural network parameterization and training, image classification and accuracy assessment, and change detection. The following sections document these procedures.

5.2.1 Image preprocessing
Each individual Landsat image in Table 3.1 was radiometrically and geometrically corrected by using the Level 1 Product Generation System (LPGS) at the USGS EROS Data Center. The pixel size was standardized to 15 meters by using the nearest-neighbor resampling strategy. As stated in Chapter Three, two adjacent scenes acquired at the same date were mosaicked to cover the entire study area. The mosaics of these scenes were further geo-referenced to the 3° Gauss-Kruger projection (117E) and the Beijing 1954 datum. As a result, the image data had the same spatial coordinate system and projection as the reference maps. Next, the TM image acquired in 2009 was geo-corrected to the ETM+ image acquired in 2000 given the latter’s high-resolution panchromatic band. Finally, these images were clipped using the administrative boundary of the Beijing metropolitan area. The resultant subsets were prepared for image classification and change detection. Figure 5.1 illustrates the specific image preprocessing procedures.

5.2.2 Land cover classification scheme

Through GPS-guided field surveys, I obtained the first-hand information about the natural and cultural landscapes in my study area. These field observations, coupled with high-resolution aerial and space-borne imagery (see Chapter Three), were utilized to design a land cover classification scheme and collect training and test samples. Finally, a mixed Anderson Level I/II land use/cover classification scheme was established (Anderson et al., 1976). This system includes five major classes: urban land, barren land, cropland/grassland, forest, and water (Yang, 2002; Yang and Lo, 2002). Table 5.1 provides a description for each class.

5.2.3 Artificial neural network parameterization and training

According to the pilot research I conducted on the use of artificial neural networks (see Chapter Four), multi-layer-perceptron feed-forward back-propagation neural networks were adopted for image classification here. Only one hidden layer was used, and the logistic-sigmoid function with training threshold 0 was used as the activation function. Input data were the seven TM image bands for 2009 and eight ETM+ image bands for 2000. Outputs were five classes described in Table 5.1. Considering its easiness of implementation and robust performance in land cover classification, the gradient descent with momentum algorithm was utilized for neural network training. Training parameters were configured as following: learning rate = 0.01;
Collect Landsat images

Standardize pixel size to 15m

Mosaic adjacent scenes

Re-project the mosaic of images

Image registration

Subset images

Images ready for classification

Figure 5.1 Working flow chart for image preprocessing.
Table 5.1 Land cover classification scheme.

<table>
<thead>
<tr>
<th>No.</th>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Urban land</td>
<td>Greater than 50 percent impervious surface; typically commercial or industrial buildings, open transportation facilities, residential areas, local road, and small open spaces with a certain amount of vegetation cover (up to 20 percent)</td>
</tr>
<tr>
<td>2</td>
<td>Barren land</td>
<td>Mainly non-impervious areas with sparse vegetation cover (less than 20 percent), such as clear-cuts, quarries, barren soil or rocks, and construction areas</td>
</tr>
<tr>
<td>3</td>
<td>Cropland/grassland</td>
<td>Crop fields, pasture, as well as cultured grasses (such as golf course, lawns and city parks)</td>
</tr>
<tr>
<td>4</td>
<td>Forest</td>
<td>Deciduous, coniferous and mixed forests and orchards (above 90 percent)</td>
</tr>
<tr>
<td>5</td>
<td>Water</td>
<td>Open water, generally with 95 percent water surface, including rivers, streams, reservoirs and lakes</td>
</tr>
</tbody>
</table>
momentum = 0.8; iterations = 1300; training goal = 0.1 in RMSE; and minimum gradient = 1e-006. The training process stopped when either the training goal or the minimum gradient was reached.

Based upon prior studies, training samples should be at least 10 times more than input neurons (Kavzoglu and Mather, 2003). Classification accuracy gradually increases as the training set size increases and the increase rate declines as the training set size gets larger (Zhuang et al., 1994; Foody and Arora, 1997; Kavzoglu and Mather, 2003). In this study, I carefully collected two independent training datasets. One was for classifying the 2000 ETM+ image; the other was for the 2009 TM image. Each dataset contains 250 samples for each land cover class. Note that the development of each dataset involved the use of the high-resolution reference images discussed in Chapter Three. These training datasets were used to train the aforementioned neural networks model.

5.2.4 Image classification and accuracy assessment

Neural networks training led to the creation of two trained models that were further applied to classify the 2000 ETM+ image and the 2009 TM image into five land cover classes (see Table 5.1). And the 2000 and 2009 land cover maps for the Beijing metropolitan area were produced.

Next, a GIS-based post-classification procedure was conducted to improve the accuracies of these two land cover maps. Specifically, sieving and majority analysis functions were run to replace isolated pixels by the majority of the surrounding landscapes. Here patches with five or fewer pixels (1125m²) were defined as isolated, and the urban-land class was excluded from the analysis because small-size urban lands were common in the study area.

After the post-classification procedure, the land cover map accuracies were assessed using the confusion matrix method that has been discussed in Section 4.3.4. Test samples were collected by using the stratified random sampling scheme and the reference data (See Chapter Three). Test datasets contain about 80 samples for each land-cover class. Confusion matrixes were computed for the 2000 and 2009 land cover maps, respectively, by using the relevant test dataset (Tables 5.2-5.3). Besides confusion matrixes, Figure 5.2 visually illustrates the accuracy of the 2009 land cover map by using an urban fringe area in the eastern Beijing as an example.
Table 5.2  Confusion matrix and Kappa statistics for the 2000 land cover map.

<table>
<thead>
<tr>
<th>Classified map</th>
<th>Reference (Pixels)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban land</td>
<td>Barren land</td>
</tr>
<tr>
<td>Urban land</td>
<td>66</td>
<td>19</td>
</tr>
<tr>
<td>Barren land</td>
<td>15</td>
<td>62</td>
</tr>
<tr>
<td>Cropland/ grassland</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Forest</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>83</td>
<td>81</td>
</tr>
<tr>
<td>Producer’s Accuracy</td>
<td>79.52%</td>
<td>76.54%</td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappa Coefficient</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.3  Confusion matrix and Kappa statistics for the 2009 land use/cover map.

<table>
<thead>
<tr>
<th>Classified map</th>
<th>Reference (Pixels)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>User’s Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban land</td>
<td>Barren land</td>
<td>Cropland/grassland</td>
<td>Forest</td>
<td>Water</td>
<td>Total</td>
</tr>
<tr>
<td>Urban land</td>
<td>64</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>79</td>
</tr>
<tr>
<td>Barren land</td>
<td>12</td>
<td>62</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>Cropland/ grassland</td>
<td>3</td>
<td>2</td>
<td>62</td>
<td>8</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>70</td>
<td>1</td>
<td>81</td>
</tr>
<tr>
<td>Water</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>78</td>
<td>84</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>400</td>
</tr>
<tr>
<td>Producer’s Accuracy</td>
<td>80%</td>
<td>77.50%</td>
<td>77.5%</td>
<td>87.5%</td>
<td>97.5%</td>
<td></td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>84.00%</td>
</tr>
<tr>
<td>Kappa Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.80</td>
</tr>
</tbody>
</table>
Figure 5.2  Visualizing the classification accuracies by comparing the image data with the classified land cover map: (a) A subset of the 2009 Landsat TM image at the eastern urban fringe (band 4-red; band 3-green; band 2-blue); (b) The classified land cover map at the same location (red-urban use; green-cropland/grassland; blue-water; dark green-forest; light grey-bareland).
To characterize urban growth, the urban land class was extracted from each of the two land cover maps, and two urban extent maps were created, which contain the urban land and non-urban land classes. These two urban extent maps were further refined through manual interpretation, and some misclassified bare lands were removed. Figure 5.3 displays the final urban extent maps for 2000 and 2009, and clearly, urban land has rapidly expanded during the period of 2000-2009.

5.2.5 Change detection

An urban land change map was produced from the 2000 and 2009 urban extent maps using the GIS overlay analysis. Urban lands here were defined as areas with greater than 50 percent impervious surface (See Table 5.1). They are, typically, commercial or industrial buildings, open transportation facilities, residential areas, local road, and small open spaces with a certain amount of vegetation cover (up to 20 percent). Figure 5.4 shows the location of the newly developed urban land during the study period. Clearly, the urban expansion in the Beijing metropolitan area largely occurred evenly at all directions, except the western suburban area. The exception may be partially accounted for by the biophysical conditions and land use policies. As discussed in Chapter Three, the study area is composed of alluvial plains in the southern and eastern parts, and hills/mountains in the northern, northwestern, and western parts. By integrating digital elevation model (DEM) data and other ancillary data, it is found that urban lands had almost approached the hills/mountains in the western part since 2000, and thus limited lands were available for any further development. In addition, the Mentougou district in the west has been zoned as the Ecological Conservation Region in the “Beijing’s regional development planning at the Eleventh Five-Year period”, which also prevented lands from being developed.

The so-called node-area development pattern can also be observed from Figure 5.4. The newly developed lands were concentrated around the urban core and sub-urban centers such as Tongzhou, Daxing and Changpin. As the expansion of these sub-centers, newly developed urban lands gradually joined the urban core and formed a continuous urban space.

Besides the urban land change map, statistical analysis was conducted to quantitatively measure the amount and rate of urban spatial growth at the metropolitan and functional zone levels, respectively (Table 5.4 and Figure 5.5). At the metropolitan level, the amount and rate of urban
Figure 5.3  The 2000 and 2009 urban extent maps.
Figure 5.4  Urban spatial growth during 2000-2009 in the Beijing metropolitan area.
Table 5.4  Statistics of the urban growth in the Beijing metropolitan area during 2000-2009 at metropolitan and functional zone levels, respectively.

<table>
<thead>
<tr>
<th>Functional Zones *</th>
<th>Districts</th>
<th>2000</th>
<th>2009</th>
<th>Amount of urban growth (km²)</th>
<th>Rate of urban growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban core</td>
<td>Dongcheng, Xicheng, Xuanwu, Chongwen</td>
<td>88.9</td>
<td>96.2%</td>
<td>88.3</td>
<td>95.6%</td>
</tr>
<tr>
<td>Extensive urban</td>
<td>Haidian, Chaoyang, Fengtai, Shijinshan</td>
<td>575.5</td>
<td>45.3%</td>
<td>864.7</td>
<td>68.0%</td>
</tr>
<tr>
<td>New urban</td>
<td>Changpin, Shunyi, Tongzhou, Fangshan, Daxin</td>
<td>679.3</td>
<td>10.7%</td>
<td>1549.3</td>
<td>24.4%</td>
</tr>
<tr>
<td>Ecological conservation</td>
<td>Yanqing, Miyun, Huairou, Pinggu, Mentougou</td>
<td>180.9</td>
<td>2.1%</td>
<td>373.0</td>
<td>4.3%</td>
</tr>
<tr>
<td>Metropolitan</td>
<td>All 18 districts</td>
<td>1524.6</td>
<td>9.3%</td>
<td>2875.3</td>
<td>17.6%</td>
</tr>
</tbody>
</table>

* The four functional zones were defined by the “Beijing regional development planning at the Eleventh Five-Year period”.

** The urban land percentage is defined as the ratio between the urban land area and the total area.
Figure 5.5  Trends of the urban growth at various functional zones from 2000 to 2009.
growth from 2000 to 2009 were 1350.7km$^2$ and 88.6%, respectively. At the functional zone level, the fastest growing area was the new urban zone with an addition of 870.0km$^2$ and a growth rate of 128.1%. In addition, urban land percentage (the urban land area/the total area) was computed because this metric can provide useful information about different urban growth modes, such as high-density urban expansion from urban centers versus low-density decentralized urban spread without a well-defined center (e.g. urban sprawl). The extensive urban zone had the largest increase in terms of urban land percentage that increased from 45.3% in 2000 to 68% in 2009.

Based on the urban land change map and statistical summaries at the metropolitan and functional zones levels, the urban growth in the Beijing metropolitan area can be summarized below:

1) Overall, the study area had experienced rapid urban growth. Urban lands had almost doubled in terms of physical area.

2) Geographically, the urban growth progressed in a mode of high-density expansion from the existing urban core toward the northern, eastern, and southern suburban areas. Distance to the urban core was a crucial factor affecting urban spatial growth. The extensive urban zone adjacent to the urban core had a dramatic growth even through its urban land area had already occupied 45.3% of the total land area in 2000.

3) The suburban centers played an increasingly important role in the urban growth, which may be attributed to two major factors. Firstly, the construction of commuting trains, highways, and other public transportation facilities in the suburban areas had greatly increased the accessibility. Secondly, abundant land supplies and the cost-effectiveness at the suburban centers encouraged housing and business development to move to the suburban centers. As a result, the new urban zone had the fastest urban growth during the period of 2000-2009.

5.3 Urban growth characterization

5.3.1 Introduction

The general trends of the urban spatial growth in the Beijing metropolitan area were investigated by the urban land change mapping and statistic analysis that have been discussed in Section 5.2. Here, a further analysis has been conducted in order to develop an in-depth understanding of
urban growth patterns and processes. This analysis was based on the use of landscape metrics because some prior studies have indicated that they can be useful for understanding urban growth patterns (e.g., Herold et al., 2002; Wu et al., 2006; Peng, et al., 2010; Shrestha et al., 2012).

This section will discuss my urban growth pattern analysis by using landscape metrics. And it will further discuss the roles of three major urban growth processes in shaping urban growth patterns, including urban compaction, urban aggregation and urban dispersion. Urban compaction refers to the compaction of existing urban lands and the re-use of previously developed lands, such as old city renovation (Pauleit and Golding, 2005). It can reduce the urban land patch number, increase the patch size, simplify the urban patch shape, and increase the landscape isolation. Urban aggregation corresponds to patch clustering to form large patches (Aguilera, 2011). It can evidently reduce the total number of patches (NP), thus increasing the patch size (Aguilera, 2011). Urban dispersion is a decentralized urban growth process by which newly developed lands disperse in the suburb areas and are away from previously developed lands. The dispersion process can increase the number of patches and landscape complexity.

A variety of GIS software packages provide functions for landscape metrics analysis, such as ArcGIS, GRASS, FRAGSTATS developed by McGarigal et al. at Oregon State University, and APACK by Mladenoff et al. at the University of Wisconsin-Madison. FRAGSTATS 3.3 was adopted in this research because of its powerful functions and the popularity.

5.3.2 Selection of landscape metrics

As stated in Chapter Two, landscape metrics quantitatively measure landscape patterns in terms of landscape composition, patch properties and spatial configuration. There are numerous landscape metrics available. Many of them are, however, correlated with one another (Riitters et al., 1995). The selection of appropriate landscape metrics is critical to understand landscape patterns (Peng et al., 2010).

In this research, two criteria were used for selecting landscape metrics. Firstly, those metrics explicitly designed for landscape ecology were excluded. For instances, the Core Area metrics measure the patterns that are related to the depth-of-edge distance from the patch perimeter. These patterns may be important for the survivor of some species, but are less indicative to urban
growth processes. Secondly, for each individual landscape aspect, only one or two typical metrics were used to avoid duplication. For example, the aggregation index (AI) was exclusively chosen to measure landscape aggregation. As a result, six landscape metrics were selected at the class level to measure landscape patterns in terms of landscape fragmentation, landscape complexity, landscape isolation, or landscape aggregation (Table 5.5). These four aspects were considered because they are useful to characterize urban growth processes, including urban aggregation, compaction, and dispersion (Aguilera et al., 2011).

5.3.3 Landscape metrics analysis

Scale plays an important role in landscape metrics analysis (Turner et al., 1989; Wickham and Riiters, 1995; O’Neill et al., 1996; Wu, 2004; Saura, 2004; Buyantuyev and Wu, 2007). For remote sensing-based landscape metrics analysis, many factors may influence the performance of landscape metrics analysis, such as the spatial resolution of remote sensor data, image classification accuracies, the pixel size of produced urban extent maps and the map extent. In this research, the spatial resolution of remote sensor data was standardized to 15 meters and image classification accuracies were evaluated using confusion matrix (see Tables 5.2-5.3). I further conducted landscape metrics analysis with various cell sizes and map extents to explore the impacts of the cell size and map extent on landscape metrics analysis. I then discussed urban growth patterns and processes at the metropolitan and functional zone levels.

To evaluate the effect of cell size, the 2009 urban extent map was used for urban pattern analysis by using landscape metrics. In the data preparation, I first recoded the urban extent map by labeling the urban land class as “1” and other land types as “2”. Then, I resampled the recoded maps with the cell sizes 30m, 50m and 100m, respectively. I chose the cell sizes between 30m and 100m because of two reasons. Firstly, the satellite images used in this research for urban land change mapping had a 30m spatial resolution, and thus cell sizes smaller than the image spatial resolution (30m) were meaningless. On the other hand, when a cell size is too large, it is hard to preserve local urban spatial patterns given the size diversity of urban lands. The resampling method adopted here was the “Majority Resampling” technique that determines the new value of the cell based on the most popular values within the filter window. With the resampled maps, I further converted them into the ASCII format and input the three ASCII files
Table 5.5  Six selected metrics for landscape pattern analysis.

<table>
<thead>
<tr>
<th>No.</th>
<th>Metrics</th>
<th>Abbreviation</th>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of Patches</td>
<td>NP</td>
<td>Total amount of patches</td>
<td>A basic measure of landscape fragmentation. A fragmented landscape has a relatively high NP value.</td>
</tr>
<tr>
<td>2</td>
<td>Patch Mean Area</td>
<td>PMA</td>
<td>Total area of patches/total amount of patches (ha)</td>
<td>A measure of landscape fragmentation by computing the mean area of patches within a certain extent. A large PMA value indicates less fragmentation.</td>
</tr>
<tr>
<td>3</td>
<td>Mean Perimeter-Area Ratio</td>
<td>MPAR</td>
<td>Average(patch perimeter/patch area)</td>
<td>A basic measure of shape complexity calculated by the mean of patch perimeter-area ratios. The decrease of shape complexity may lower the MPAR value, but the increase of the patch size can also cause the decrease of the MPAR</td>
</tr>
<tr>
<td>4</td>
<td>Landscape Shape Index</td>
<td>LSI</td>
<td>Total perimeter length of all patches/possible minimum perimeter length with the same area</td>
<td>A measure of shape complexity. The simpler the shape is, the lower the LSI value.</td>
</tr>
<tr>
<td>5</td>
<td>Mean-Euclidean Nearest Neighbor Distance</td>
<td>M-ENN</td>
<td>Average(edge-to-edge distance from patches to their nearest neighbor) (meter)</td>
<td>A measure of landscape isolation. A high M-ENN value indicate the great landscape isolation</td>
</tr>
<tr>
<td>6</td>
<td>Aggregation Index</td>
<td>AI</td>
<td>Number of like adjacencies/possible maximum number of like adjacencies</td>
<td>A measure of landscape aggregation. AI ranges between 0 and 1. AI reaches the value of 1 when the class is maximally clumped into a single, compact</td>
</tr>
</tbody>
</table>
into FRAGSTAT 3.3 for the landscape metrics analysis. The six selected landscape metrics were computed from the maps with various cell sizes.

To analyze urban growth patterns at the metropolitan and function zone levels, I used both the 2000 and 2009 urban extent maps with the cell size 30m. I first clipped the two maps using the boundaries of the four functional zones: urban core, extensive urban, new urban and ecological conservation zones (see Table 5.4). As a result, eight zone maps were produced. Then, I prepared these zone maps and the metropolitan maps through recoding, resampling and conversion procedures. Next, I conducted landscape metrics analysis for each individual zone and the metropolis for 2000 and 2009, respectively, based on the relevant urban extent map. Consequently, the six selective landscape metrics were computed from the 2000 and 2009 metropolitan and functional zone maps.

Finally, I evaluated the scale dependency of the landscape metrics analysis by comparing landscape metrics values with various cell sizes and map extents. Then, I discussed urban growth pattern in terms of landscape fragmentation, shape complexity, landscape isolation, and landscape aggregation at the metropolitan and functional zone levels. Due to the interrelations between urban growth patterns and urban growth processes, I further discussed the roles of three major urban growth processes in shaping the observed urban growth patterns.

5.3.4 Results and discussion

5.3.4.1 Scale dependency: cell size and zoning

The experiments with various cell sizes indicate that landscape metrics are very sensitive to the cell size. From Table 5.6 and Figure 5.6, the values of the metrics, namely, NP, LSI, MPAR and AI, increased greatly as the cell size decreased from 100m to 30m. With a smaller cell size, more patches may be detected, as indicated by NP, more complex patches shape may be observed, as suggested by LSI and MPAR, and higher landscape aggregation may be measured based on AI. In addition, two metrics, namely, PMA and M-ENN, show a positive correlation with the cell size. As the cell size decreases, patches area may be reduced and the average distance between two nearest patches decreases. These variations suggest that remote sensing-derived landscape metrics analysis with different cell sizes can lead to a misunderstanding of urban growth
Table 5.6  Results of the landscape metrics analysis with cell sizes of 30m, 50m, and 100m, respectively, based on the 2009 urban extent map.

<table>
<thead>
<tr>
<th>Cell size (m)</th>
<th>NP*</th>
<th>PMA</th>
<th>MPAR</th>
<th>LSI</th>
<th>M-ENN</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5921</td>
<td>50</td>
<td>552</td>
<td>107</td>
<td>217</td>
<td>94.1</td>
</tr>
<tr>
<td>50</td>
<td>5326</td>
<td>55</td>
<td>450</td>
<td>103</td>
<td>261</td>
<td>90.6</td>
</tr>
<tr>
<td>100</td>
<td>2856</td>
<td>76</td>
<td>259</td>
<td>88</td>
<td>390</td>
<td>83.9</td>
</tr>
</tbody>
</table>

* See full landscape metrics names in Table 5.5.
Figure 5.6  Trends of landscape metrics in 2009, namely, NP, PMA, LSI, MPAR, M-ENN and AI, as related to various cell sizes. Full names of the metrics are given in Table 5.5.
Therefore, it is important to standardize the base map cell size before a landscape metrics analysis can be conducted.

Figures 5.7-5.8 and Table 5.7 show the landscape metrics values at various map extents: the metropolis, urban core, extensive urban, new urban and ecological conservation zones. Clearly, the absolute landscape metrics values vary greatly as the map extent changes. Compared to the landscape metrics analysis at the metropolitan level, spatial variations of urban patterns were detected at the functional zone level. Also, due to the sensitivity of landscape metrics analysis to the map extent, it is necessary to use the same map extent for examining urban pattern changes over time.

To summarize, both the cell size and the map extent can have a significant influence on landscape metrics analysis. Therefore, we need to be careful when interpreting landscape metrics results, particularly when comparing their absolute values. It is meaningless to compare urban patterns quantified by the landscape metrics analysis using base maps with different cell sizes. Also, it is necessary to conduct the landscape metrics analysis at the same map extent for quantifying urban pattern changes over time.

5.3.4.2 Urban pattern-process analysis at the metropolitan level

As stated in Section 5.3.3, a set of experiments were conducted to measure the 2000 and 2009 urban patterns with an identical cell size of 30m at the metropolitan level (Table 5.7). The six selective landscape metrics were computed to measure urban pattern changes during the period of 2000-2009 from four aspects: landscape fragmentation, shape complexity, landscape isolation, and landscape aggregation, as discussed in Section 5.3.2. By comparing the experimental results for 2000 and 2009, the urban growth pattern in the Beijing metropolitan area can be characterized by less fragmented, more aggregated, higher shape complexity and increased isolation.

Firstly, the urban landscape as a whole in 2009 was less fragmented than 2000, as indicated by the values of NP and PMA (see Table 5.7). Experimental results show that the NP value decreased while the PMA value increased from 21.1 Ha to 49.5 Ha (Table 5.7). Secondly, LSI and MPAR that measure patch shape complexity show opposite trends. Specifically, the slight
Figure 5.7  Results of the landscape metrics analysis at the functional zone level in 2000. Full names of the metrics are given in Table 5.5.
Figure 5.8  Results of the landscape metrics analysis at the functional zone level in 2009. Full names of the metrics are given in Table 5.5.
Table 5.7  Results of the landscape metrics analysis at the metropolitan level in 2000 and 2009.

<table>
<thead>
<tr>
<th>Years</th>
<th>NP*</th>
<th>PMA</th>
<th>MPAR</th>
<th>LSI</th>
<th>M-ENN</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>7339</td>
<td>21.1</td>
<td>631.8</td>
<td>92.5</td>
<td>210.1</td>
<td>93</td>
</tr>
<tr>
<td>2009</td>
<td>5921</td>
<td>49.5</td>
<td>522.4</td>
<td>106.9</td>
<td>216.8</td>
<td>94.1</td>
</tr>
</tbody>
</table>

* See full landscape metrics names at Table 5.5.
LSI increase suggests the increase of shape complexity, but the MPAR decline indicates less shape complexity. Because the increase of patch size may also lead to the MPAR decline and the patch mean area (PMA) doubled in 2009, the drop of MPAR is highly likely to be attributed to the increase of patch size, rather than the decrease of shape complexity. Therefore, the LSI analysis result was used to characterize the landscape shape complexity. Thirdly, compared to 2000, urban lands in 2009 were more aggregated but with an increasing distance to their neighbors, as indicated by the increase of AI and M-ENN.

The observed urban pattern changes suggest that urban compaction and aggregation processes had occurred in the Beijing metropolitan area during the study period. The much less fragmented and more aggregated landscape suggests the effects of the urban compaction and aggregation processes. The slightly increased landscape isolation could be the result of urban compaction, such as old city renovation. Urban dispersion may also have occurred in the study area. However, the importance of these urban growth processes may vary over space and over time. It is necessary to characterize urban growth at the functional zone or even larger scale in order to understand the spatial and temporal variations of urban growth patterns and processes.

5.3.4.3 Urban pattern-process analysis at the functional zone level

A set of experiments were also conducted to measure the 2000 and 2009 urban patterns with an identical cell size of 30m at the functional zone level (Figures 5.7-5.8). By comparing the experimental results for 2000 and 2009 at different functional zones, the urban growth patterns in the study area and the underlying processes were measured and discussed here.

In general, landscape became less fragmented and more aggregated in all zones, except the ecological conservation zone. The urban patterns in the extensive urban zone had experienced the largest change, which tended to resemble those in the urban core. Patterns in the urban core changed the least.

From Figure 5.7, the urban core was characterized by the high-density urban landscape with extremely low fragmentation measured by the NP and PMA, and high landscape aggregation measured by the AI in 2000, which was quite different from the patterns in other zones. During the study period, landscape fragmentation and aggregation didn’t change much while the shape
complexity decreased and landscape isolation increased, as indicated by the MPAR and M-ENN, respectively (Figure 5.9). In 2000, like other two zones, the extensive urban zone was much different from the urban core in terms of landscape fragmentation, shape complexity, landscape isolation and landscape aggregation (Figure 5.7). However, as illustrated in Figure 5.8, the urban patterns in the extensive urban zone quite resembled to those in the urban core during the period of 2000-2009. Specifically, the landscape fragmentation plunged while urban lands became highly aggregated, as indicated by the NP, PMA, and AI measurements (Figure 5.9). Also, the landscape complexity dropped greatly and urban lands became isolated in this zone, as suggested by the LSI, MPAR and M-ENN (Figure 5.9). Considering the adjacent locations of the extensive urban zone and the urban core, the urban center in the Beijing metropolitan area had clearly experienced a high-density expansion.

Figure 5.9 shows that urban lands were gradually aggregated and became less fragmented in the new urban zone, as measured by the AI, NP and PMA values, which show the same trend with that in the extensive urban zone. But the shape complexity increased greatly and urban landscapes became less isolated, as indicated by the LSI and M-ENN, respectively, which was opposite to the urban core and extensive urban zones. Different from other three zones, landscape fragmentation increased and the aggregation process was quite limited in the ecological conservation zone, as indicated by the increased NP, the barely changed PMA and the same AI for the period of 2000 to 2009 (Figure 5.9).

The roles of the major urban growth processes can be analyzed based on the observed urban growth patterns, including urban compaction, urban aggregation and urban dispersions. These three urban growth processes was discussed in Section 5.3.1. The urban core was characterized by the barely changed extremely low landscape fragmentation and high aggregation, and the reduced shape complexity and increased landscape isolation suggest that urban compaction, such as old city renovation and urban greening work, dominated the urban growth in this area during the period of 2000-2009. This urban growth process re-developed the existing complex urban lands into well-organized intensive urban use.

In the extensive urban zone, urban land percentage increased rapidly (see Figure 5.5), the landscape fragmentation plunged, and landscape aggregation increased greatly (see Figure 5.9).
Figure 5.9  Trends of urban growth patterns at different functional zones during 2000-2009, indicated by landscape metrics analysis. Full names of the metrics are given in Table 5.5.
These urban growth patterns can be explained by both urban compaction and aggregation processes. The reduced landscape complexity and increased landscape isolation further justify the dominance of the high-density re-development of existing urban lands (urban compaction) and the clustering of newly developed lands (urban aggregation) in this zone.

Different from the urban core and extensive urban zones, the urban growth in the new urban zone was characterized by urban aggregation and dispersion processes. On one hand, some newly developed land clustered around the sub-urban centers, producing the less fragmented and slightly aggregated landscape patterns. On the other hand, the urban dispersion process created new discrete urban lands, which increased landscape complexity and fragmentation and reduced the isolation and aggregation level. The observed urban growth patterns in this area were the compositive results of these two opposite processes. The ecological conservation zone was dominated by the urban dispersion process given the greatly increased landscape fragmentation, reduced landscape isolation, and barely changed aggregation level. Finally, urban growth analysis at a larger scale may provide more information about the spatial and temporal variations of urban growth patterns and processes.

5.3.5 Summary

In Section 5.2, I analyzed urban growth characteristics in the Beijing metropolitan area by geo-visualizing urban land change and statistically summarizing urban land areas and urban land percentages at the metropolitan and functional zone levels for 2000 and 2009. In this section, I quantitatively analyzed urban growth patterns using landscape metrics analysis and discussed urban growth processes at the metropolitan and functional zone levels. To summarize, through the landscape metrics analysis, urban growth patterns and processes in the Beijing metropolitan area during the period of 2000-2009 were characterized by the following features:

1) The high-density urban center extended greatly toward the extensive urban zone, which led to substantial urban pattern changes there.

2) Urban compaction process (such as urban reconstruction/renovation and urban greening work) dominated the urban core and part of the extensive urban zone, which produced the observed landscape pattern changes in the areas.
3) Urban aggregation process greatly changed the landscape patterns in the two zones adjacent to the urban core, particularly the extensive urban zone. It reduced landscape fragmentation and increased landscape aggregation.

4) Urban growth in the areas away from the urban core, such as the ecological conservation region, was ruled by the urban dispersion process. As a result, the number of urban land patches increased but the patch size barely changed. The landscape isolation decreased but the aggregation level remained unchanged.

5.4 Spatial and temporal dynamics of urban growth

5.4.1 Introduction

Section 5.2 discusses the general trends of urban growth through geo-visualization and statistical analysis, and Section 5.3 examines urban grow patterns and processes at the metropolitan and functional zone levels through the landscape metrics analysis. This section is designed to explore the spatial and temporal variations of urban growth patterns at the cell level through a moving windows analysis and GIS-based spatial analysis. It will also introduce the categorization of urban growth patterns in the study area and the analysis of urban spatial growth dynamics. Specifically, urban growth dynamics analysis consisted of the following analytical procedures: landscape metrics analysis with moving windows, categorizing urban patterns through a contour analysis, and spatio-temporal features of urban growth (Figure 5.10).

5.4.2 Landscape metrics analysis with moving windows

Landscape metrics analysis with moving windows (i.e., local neighborhood structure) is to compute metrics at a window with a specified size and shape (McGarigal and Marks, 1995). As the window passes over the landscape, each selected metrics will be returned to the focal cell, and a map of the metrics will be produced. This method can effectively measure and quantify the spatial variations of landscape patterns at the local level. Given a landscape metrics $m$ and a moving window size $n$, for individual cells of the input data, the moving windows analysis will calculate the metrics $m$ by using a $n \times n$ square or a circle (radius = $n/2$) that centers at the cell as the extent (see Figure 5.11). The resulting metrics map can help analyze the changes of the metrics $m$ over space.
Figure 5.10  The analytic procedures for urban growth dynamics analysis.
Figure 5.11  Illustration of the landscape metrics analysis with moving windows. A certain size window passes through cells one by one across the study area. Landscape metrics at the focal cell of these windows are then computed by using the window as the extent.
In this research, the 2000 and 2009 urban extent maps were used as the input data for the moving windows analysis. To reduce the computation complexity, the cell size of the input data was resampled to 50 meter, and the extent was the entire Beijing metropolis. The selection of the moving window size is critical to the moving windows analysis. If it is too small, the analysis may over-estimate local landscape variations and overlook some important landscape patterns operating over a large space; if too large, the analysis may neglect local spatial variations of landscape patterns. Considering that the urban core in the Beijing metropolitan area was around 20km *20km and the newly developed sub centers ranged from 3km*3km to 10km*10km, the moving window size was set to 50 pixels (2.5km in radius and 5km*5km in area). This moving window size should be sufficient for detecting urban growth patterns over space and time. The six metrics in Table 5.5, namely, NP, PMA, LSI, M_ENN, MPAR and AI, were computed by using FRAGSTATS3.3 with the above parameters setting. Figures 5.12-5.13 illustrate the results of the landscape metrics analysis with moving windows.

5.4.3 Categorizing urban growth patterns

In this section, I firstly conducted a contour analysis to the metrics maps produced by the moving windows analysis discussed in Section 5.4.2. The reason why I did a contour analysis is that the distribution of landscape metrics contours can not only pinpoint the urban center/suburban centers, but also demonstrate how the influence of these centers on their periphery declines as the distance increases. According to the contour distribution, I further delineated the urban zones that represent three typical urban growth patterns in the Beijing metropolitan area during the study period.

5.4.3.1 Contour analysis

As discussed in Section 5.4.2, six landscape metrics maps were produced for 2000 and 2009 by using the moving windows analysis (see Figure 5.12 and 5.13). These metrics are the number of patch (NP), patch mean area (PMA), landscape shape index (LSI), mean perimeter-area ratio (MPAR), mean-Euclidean nearest neighbor (M-ENN) and aggregation index (AI) (see Figure5.5). For each metrics map, a contour analysis was conducted, and different urban growth patterns over space were identified from the contour maps by using the two criteria:
Figure 5.12  Results of the landscape metrics analysis with moving windows for 2000.
Figure 5.13  Results of the landscape metrics analysis with moving windows for 2009.
1) Closeness of contours, which indicates the impacts of growth nodes/lines, such as urban center, sub-centers, major transportations, and industrial development zones; and

2) Density of closed contours, which suggests how fast landscape metrics change.

Figure 5.14 shows the contour analysis results. From the 2000 NP contour map, three urban patterns were identified using the above criteria (see Figure 5.14a). Zone 1 was enclosed by the NP contour equal to “9” (the red line in Figure 5.14a). This area was characterized with broken, sparse NP contours, which indicates the high-density urban center with little changes in landscape fragmentation. Zones 2 (the purple line in Figure 5.14a) and 3 (the green line in Figure 5.14a) were identified by the NP contour equal to “27”. In Zone 2, NP contours were dense and closed, which suggests that landscape fragmentation greatly increased as the distance to the urban center increased. Zone 3 had contours larger than “27”. The high NP values in this zone suggest a high landscape fragmentation. When using these thresholds for the 2009 map, only Zone 1 was successfully identified (Figure 5.14b), which suggests that this zone had expanded substantially after 2000. Outside of this zone NP contours were unclosed and distributed irregularly, suggesting that the landscape fragmentation in 2009 may no longer be closely associated with the distance to the urban center as the influence from suburban centers or other growth nodes emerged in this area.

The PMA contour analysis has recognized one zone from the 2000 and 2009 metrics maps by using the contour line “100 ha”, respectively (Figure 5.14c-d). The identified zone had contours greater than “100 ha” that were unclosed, dense, and distributed irregularly. The distribution of the PMA contours in this zone suggests a relatively low landscape fragmentation and no clear impacts of the urban core upon the patches size.

The LSI contour analysis has identified three patterns (see Figure 5.14e-f). Zone 1 in both 2000 and 2009 maps was determined by using the contour line equal to “6” (the red line in Figure 5.14e-f). This zone was characterized with closed, sparse contours with the LSI values smaller than “6”, which indicates that the LSI value increased slowly as the distance to the urban center increased, so did the landscape shape complexity. Zone 3 was defined by using the contour equal to “9” (the green line in Figure 5.14e-f), within which contours with the LSI values higher than “9” were locally closed around some sub-centers. Clearly, the landscape complexity in this zone was influenced by the suburban centers rather than the urban core. The area between Zone 1 and
Figure 5.14  Figure 2 Results of the contour analysis: NP, PMA, LSI and AI. Areas with different urban patterns (zones) were recognized by using the same thresholds for 2000 and 2009. Full names of the metrics are given in Table 5.5.
the inner boundary of Zone 3 was defined as Zone 2. Unlike Zones 1 and 3, the LSI in Zone 2 quickly increased as the distance to the urban center increased, but the contours were unclosed, which suggests the impacts of the urban core on the landscape complexity in this Zone.

Two patterns (zones) were identified from the AI contour maps by using the two criteria discussed above. Zone1 had closed and sparse AI contours, which implies that the aggregation level in this area had a negative correlation with the distance to the urban center. The closer to the urban center it is, the more aggregated urban lands are. But the decline of landscape aggregation was not remarkable as the distance to the urban core increased, given the low density of the AI contours in Zone1. Zone 2 had closed and dense contours where the aggregation level plunged as the distance to the urban center increased. The rest of the study area didn’t show any noticeable patterns in terms of landscape aggregation. With regard to spatial distribution of MPAR and M-ENN, no clear patterns were observed through the contour analysis.

5.4.3.2 Categorizing three urban patterns

Section 5.4.3.1 identified urban patterns in terms of each individual metrics. This section pinpointed three universal urban patterns (zones) by synthesizing the patterns recognized from those metrics using an overlay analysis. The resultant three universal urban patterns (zones) were unique in terms of landscape fragmentation, shape complexity and landscape aggregation.

From Figure 5.15a, Zone 1 identified by using the AI contours was quite consistent with the first zone from the NP and LSI contour analysis for 2000. This was also applied to the 2009 map (see Figure 5.16a). Therefore, Zone 1 of the AI contours was defined as the high-density urban zone (Universal Zone 1). The urban pattern in this zonal area was characterized by extremely low fragmentation and shape complexity as well as high aggregation. In this zone, the distance to the urban center had a somewhat impact upon the landscape aggregation and shape complexity, as indicated by the enclosed and sparse AI and LSI contours, but such an impact was marginal upon the landscape fragmentation, as suggested by the sparse and irregular NP and PMA contours.

Zone 2 from the AI contour analysis was essentially coincident with the Zones 2 from the NP and LSI contours for 2000(see Figure 5.15b). Zones 2 from the AI and LSI contours for 2009 were also consistent (see Figure 5.16b). This area was therefore defined as the fast growing zone (Universal Zone 2), which was strongly affected by the urban center. In this zonal area, as the
Figure 5.15  The three universal zones identified through an overlay analysis for 2000. (a) Relative locations of Zone 1 from NP, LSI, and AI contour analysis; (b) relative locations of Zones 2-3 from NP, LSI and AI contour analysis; 3) relative locations of universal zones 1-3.
Figure 5.16  The three universal zones identified through an overlay analysis for 2009. (a) Relative locations of Zone1 from the NP, LSI, and AI contour analysis; (b) relative locations of zones 2-3 from the LSI and AI contour analysis; 3) relative locations of the universal zones 1-3.
distance to the urban center increased, the landscape aggregation plunged, the landscape fragmentation rapidly increased, and the landscape complexity increased slowly.

Universal Zone 3 was determined by using Zone 3 from the LSI contours which was largely a subset of Zone 3 from the NP contours and adjacent to the fast growing zone (Universal Zone 2). The urban center had a limited impact upon the landscape aggregation and fragmentation in this zonal area. Instead, local centers had a larger influence on the landscape complexity. In addition, landscapes in this area were highly fragmented.

In addition, the zone identified from the PMA contours was located between Universal Zones 1 and 2, indicating that the size of urban lands in these two zones were much higher than the rest of the area. Figures 5.15c and 5.16c show the relative locations of these three Universal Zones for 2000 and 2009.

5.4.4 Spatio-temporal dynamics of urban growth

Given the complexity of urban systems, urban growth and landscape changes can be affected by numerous factors in the biophysical and socio-economic contexts as discussed in Chapter Two. Many complex behaviors are involved in the urban development, such as nonlinearity, self-organization, path-dependency, emergence, and reciprocal feedback (Batty, 2005; An et al., 2005; Cadenasso et al., 2006; Liu et al., 2007; Manson, 2008). It can be quite challenging to simulate urban growth and landscape changes through various modeling techniques. Even some advanced models, such as agent-based model or Cellular Automata model, can only simulate one or two complex behaviors, which may introduce large uncertainties to their predictions for future urban development (e.g., Theobald and Hobbs, 1998; Li and Yeh, 2001; Parker et al., 2003; Kocabas and Dragicevic, 2006; Xie et al., 2007).

The approach I proposed here attempts to analyze the urban processes from the observed urban pattern changes that are the consequences of these processes. Through the understanding of the urban processes in the study area, I intend to explore the spatial and temporal features of urban growth. This method can have some merits as it can avoid some computationally intensive simulation work and it does not need to account for numerous driving forces and their complicated interactions.
As we can see from Figure 5.17, the extent of three Universal Zones identified in Section 5.4.3 had expanded substantially during the period of 2000-2009. Based on the spatial and temporal features of the evolution of these Zones and their corresponding patterns, we can analyze the underlying urban processes and then discuss different stages of urban land development.

This method consisted of several procedures. Firstly, I compared the extents of three Universal Zones for 2000 and 2009. I found that these three Zones were associated spatially and temporally. Specifically, Zone 1 in 2000 was still in the high-density urban zone in 2009; Zone 2 in 2000 was developed into Zone 1 in 2009; and Zone 3 in 2000 was mostly transformed into Zone 1 and Zone 2 in 2009 (Figure 5.17). I concluded three urbanizing stages in the Beijing metropolitan area during the study period by exploring the spatial and temporal evolution of urban patterns. These three stages are closely associated with the identified three Universal Zones with unique urban patterns. Zone 3 was at the first stage, Zone 2 was at the second stage, and Zone 1 was at the final stage. These stages were dominated by different urban growth processes, thus producing unique patterns.

Stage 1 is the accumulating stage of urban growth in the Beijing metropolitan area. Zones 3 in 2000 and 2009 were at this stage. For the areas at this stage, urban dispersion dominates urban growth, which greatly increases landscape fragmentation and complexity through the development of a large amount of discrete urban lands. Newly developed urban lands tend to distribute close to local urban centers or growth nodes/lines, rather than the urban core. Given the dominating urban growth process, urban patterns at this stage are characterized by increased landscape fragmentation and shape complexity, barely changed low landscape aggregation, and decreased landscape isolation, as indicated by the patterns observed in Zone 3 for 2000 and 2009.

When urban lands have accumulated to a certain level, the urban aggregation process may outpace the dispersion process and transform urban patterns in a different way. Stage 2 is the fast urban growing stage in the Beijing metropolitan area. Zones 2 in 2000 and 2009 were at this stage. It is a transitional period when urban aggregation dominates urban growth that produces high-density urban patterns. For the areas at this stage, the urban core should have an important influence on the aggregation process, and therefore, urban patterns can vary greatly according to the distance to the urban center. When compared to the areas at Stage 1, landscape fragmentation
Figure 5.17  Comparisons of the recognized universal zones between 2000 and 2009. (a) Location of the 2009 high-density urban zone (Universal Zone 1) compared to three 2000 Universal Zones; (b) location of the 2009 fast growing zone (Universal Zone 2) compared to three 2000 Universal Zones; and (c) location of the 2009 Universal Zone 3 compared to the three 2000 Universal Zones.
at Stage 2 drops, landscape aggregation significantly increases, and shape complexity decreases. The shorter distance to the urban center it is, the more dramatic urban pattern changes there are.

Stage 3 is the final stage of the urban growth in the Beijing metropolitan area. Zones 1 in 2000 and 2009 were at this stage. This stage is characterized by high-density urban lands. Compared to the areas at the previous two stages, landscapes at this stage are extremely aggregated with very low fragmentation and complexity. Both urban aggregation and urban compaction processes may occur in this area, which further reduce landscape fragmentation and increase landscape isolation.

With this three-stage urban growth model, the urban pattern changes over space from 2000 to 2009 were measured by using three urbanizing stages: accumulating, fast growing and high-density well-developed. Assuming that urban growth in the Beijing metropolitan area would continue with the same development trajectory, Zone 2 in 2009 would develop toward high-density urban land, and Zone 3 in 2009 would experience fast growing in the near future. Note that the limitation of this approach relies on the assumption of the unchanged urban development mode but this method can help identify the general trends of future urban spatial growth.

5.5 Conclusions

In this chapter, I proposed three approaches for urban growth pattern characterization, and discussed urban growth processes at the metropolitan, functional zone, and cell levels. These three approaches examined urban spatial growth from different perspectives. The first approach investigated the general trends of urban growth by using the geo-visualization and statistical summaries, as discussed in Section 5.2. The second approach quantified urban growth patterns and discussed urban growth processes at the metropolitan and functional zone levels by using landscape metrics analysis, as discussed in Section 5.3. The third approach provided an in-depth understand of the spatial variations of urban growth patterns and processes at the cell level and explored the spatio-temporal urban growth dynamics using the moving windows analysis and GIS-based spatial analysis, as discussed in Section 5.4. These approaches have been successfully applied to the Beijing metropolitan area.
By integrating the analysis results from these methods, the urban growth in the Beijing metropolitan area can be summarized:

1) Mononuclear concentric growth. The three approaches came to the same conclusion that dramatic urban expansion had occurred from the urban core to the periphery. Comparing to the suburban centers, the urban core had a significant impact upon urban spatial growth.

2) High-density urban growth. The city of Beijing grew in a mode of continuing high-density expansion from the urban core, rather than the low-density decentralized urban sprawling that has been popular in many U.S. cities.

3) Three stages of urban growth. The urban growth in the Beijing metropolitan area experienced three stages with unique urban growth patterns and processes. Stage 1 was the accumulation stage dominated by urban dispersion. The urban patterns were characterized by high landscape fragmentation and complexity, low landscape aggregation, and low landscape isolation. Stage 2 was the fast growing stage influenced by urban aggregation. The urban core significantly influenced urban development at this stage. Landscapes at this stage were less fragmented and complex, and became more aggregated and isolated. Stage 3 was the high-density development stage. The areas at this stage had extremely low fragmentation and complexity and high landscape aggregation. Both urban compaction and aggregation processes influenced the urban pattern changes at this stage.

4) Growth of sub-centers. Based on the locations of the three Universal Zones in 2000 and 2009 in Figures 5.15-5.16, sub-centers had clearly developed in the study area although their importance was still behind the urban core. A multi-nuclear pattern may emerge in the future. Nevertheless, linear patterns along major transportation facilities were not been observed.

5) Effects of land use policies. The political zoning had an important influence on the urban growth direction. The extensive urban zone (Haidian, Chaoyang, Shijingshan, and Fengtai) had experienced a rapid growth during 2000-2009, along with the new urban zone (Changping, Shunyi, Tongzhou, Fangshan, and Daxin). And public resources and services (such as transportation, utilities, lands and state-owned funds) quickly moved
toward these areas, largely due to land use policy which greatly promoted the development in these areas. In addition, the establishment of industrial development zones also helped bloom the economy in the nearby areas.

6) Impacts of biophysical conditions. The urban land in the Beijing metropolitan area mainly expanded toward three directions: northern, eastern and southern. The hills/mountains at the western part inevitably constrained the development. In addition, the northern mountain areas are the sources for drinking water (such as Guanting reservoir and Miyun reservoir) and winds primarily blew from north-west to south-east. For the environmental and ecosystem protection, manufactures were gradually moving out of northern and western ecological conservation zones.
CHAPTER SIX: SUMMARY AND CONCLUSIONS

This dissertation research has aimed at the investigation of urban growth patterns, processes, and their relevance through the lens of complexity theory to improve our understanding of the spatial and temporal dynamics of urban growth in a rapidly growing metropolitan area. An improved research methodology has been developed, which centers on urban pattern-process analysis with support of a technological framework of integrating artificial neural networks (ANNs) and GIS technologies with satellite imagery processing. The research objectives discussed in Section 1.2 have been addressed through the research efforts with two components: 1) examining the use of artificial neural networks for improving land cover classification from remote sensor data (see Chapter Four); and 2) characterizing urban spatial growth at the metropolitan, functional zone, and cell levels by using different methods (see Chapter Five). Center to this research are remote sensor data from which urban land changes have been detected and mapped by ANNs. Due to the spatial and temporal heterogeneity and scale multiplicity of urban growth, three approaches, namely, urban land change mapping, landscape metrics analysis, and moving windows analysis, have been adopted to characterize urban growth dynamics from different perspectives at different scales. The urban growth features of a fast growing metropolitan area at the information age have been studied by using the Beijing metropolitan area as a case study area. This dissertation contrasts with past works on urban land changes that have focused primarily on urban land change detection from remote sensor data without providing in-depth insight into urban spatial growth patterns and the relevant urban growth processes.

The first part of this dissertation research has been to investigate the performance of artificial neural networks for improving land cover classification from remote sensor data (see Chapter Four). The use of artificial neural networks has been examined from three aspects: neural networks architectures, training algorithms, and internal parameters settings. A set of algorithmic and non-algorithmic parameters of neural networks have been evaluated in terms of their sensitivity to image classification performance by using trial-and-error approach. As a result, a systematic approach that can guide the use of artificial neural networks has been proposed. There are some limitations of my research on the use of ANNs for image classification. In this research, the experiments were conducted using a Landsat ETM+ image in the urban environment. The
neural networks, training algorithms, and internal parameters settings were then evaluated based on their performance in urban land classification from Landsat imagery. The optimal ANNs configuration concluded from this research may not be applicable to image classification in the non-urban environment. In addition, further research is needed to test whether the optimal ANNs configuration for Landsat images can also lead to the best performance in land cover classification from other remote sensor data.

The second part of this dissertation research has been to detect and visualize urban land change from remote sensor data by using the Beijing metropolis as a case study area (see Section 5.2). The optimized artificial neural networks in Chapter Four have been used to derive land cover information from two Landsat images acquired in 2000 and 2009, respectively. A map-to-map comparison method has been adopted to detect urban landscape changes. The general trends of urban growth have been further analyzed based upon the urban land change map. Results revealed that urban lands in this area had almost doubled during the period of 2000-2009 (increasing rate=88.6%), and urban expansion had largely been occurring evenly at all directions, except the western suburban area where the biophysical conditions (e.g., mountains/hills) and land use policies had constrained the expansion. Through visualizing and statistically summarizing urban land changes, urban growth in the Beijing metropolitan area can be characterized by high-density spread from the urban core to rural areas during the study period.

The third part of this dissertation has been to characterize urban growth patterns and processes at the metropolitan and functional zone levels by using the landscape metrics analysis (see Section 5.3). Six landscape metrics have been selected carefully to measure landscape patterns in terms of landscape fragmentation, shape complexity, landscape isolation, and landscape aggregation. The scale dependency of urban growth patterns has been analyzed from two aspects: cell and extent. The importance of three major urban growth processes (e.g., compaction, aggregation, and dispersion) in the evolution of urban form have been discussed based on the observed patterns. It is found that the high-density urban center had extended greatly toward the extensive urban zone, and that urban aggregation process had dominated the two zones adjacent to the urban core while urban dispersion process had greatly influenced the pattern changes in the outer suburb area.
The last part of this dissertation has been to characterize urban spatial growth at the cell level by using the moving windows analysis and GIS-based spatial analysis (see Section 5.4). Landscape metrics analysis with a 5*5km² moving window has been used to measure spatial and temporal variations of urban growth and to provide remarkable insight into urban growth dynamics. Six landscape metrics maps have been produced, showing the changes of urban patterns over space. Three urbanizing stages and corresponding urban patterns (e.g., Urbanizing Zones) have been recognized by spatially and temporally analyzing the landscape metrics maps using GIS-based spatial analyses. Stage 1 is the accumulating stage when urban dispersion dominates urban growth and produces fragmented and complex landscape; Stage 2 is the fast urban growing stage when urban aggregation characterizes urban growth and produces high-density urban patterns; and Stage 3 is the well-developed stage characterized by high-density urban use. Through exploring the spatial and temporal development of urban patterns, this dissertation has found that urban lands at different urbanizing stages in the Beijing metropolitan area have had different urban forms.

Finally, urban growth features in the Beijing metropolitan area have been summarized by synthesizing the findings of the three approaches: urban land change mapping, landscape metrics analysis and moving windows analysis at the metropolitan, functional zone, and cell levels. Briefly, the Beijing metropolitan area had experienced mononuclear concentric high-density growth with the emergence of some sub-centers during the period of 2000-2009, and biophysical conditions and land use policies had played an important role in shaping the urban form evolution.

The dissertation contributes to the literature from conceptual, technological, and application perspectives. At the conceptual level, this dissertation research has explored urban form, urban process and their relevance in a large complex metropolitan area which has been undergoing rapid changes driven by population explosion and accelerated economic growth. This exploration has provided insights into urban growth dynamics in particular urban societies that are not comparable to either industrial or post-industrial cities in the United States and that have unique urban form and development trajectory due to technological robustness and contemporary international and domestic socio-economic conditions. In addition, this dissertation has revealed
the temporal properties of urban growth patterns and processes, which can be used to better understand urban growth dynamics without conducting driving force analysis and computationally intensive simulations.

Technologically, this project has developed an improved methodology for imagery-based urban growth characterization that combines artificial neural networks, remote sensing, and GIS-based spatial analysis for mutual reinforcement of the utility of these techniques. Firstly, a systematic approach that can guide the use of artificial neural networks in satellite image classification has been developed based on intensive experiments. Secondly, three methods, including GIS-based urban land mapping, landscape metrics analysis, and moving window analysis, have been used to characterize urban growth patterns and processes at the metropolitan, functional zone, and cell levels. Thirdly, spatial and temporal features of urban growth have been discussed to better understand urban growth dynamics.

At the application level, this dissertation has established a well-documented study of urban spatial growth in the Beijing metropolitan area. The project has mapped historical urban land extents in 2000 and 2009. It has also measured urban growth patterns and analyzed the related urban growth processes at the macro- and micro-scales. It has further pinpointed three universal urban patterns (e.g., Urban Zones) and revealed three urbanizing stages based on the observed urban pattern changes. These can be useful for academic research on urban systems and urban land use planning and resources and environmental management for sustainable development. Most importantly, the conceptual and technological frameworks proposed in this dissertation can be applicable to urban spatial growth characteristics for other large complex metropolis.
REFERENCES


BIOGRAPHICAL SKETCH

Libin Zhou was born in China. She graduated with a Bachelor of Science degree in Physical Geography from Nanjing University, China in 2000. Three years later, she obtained a Master of Science degree in Cartography and Geographic Information Science from the same institution. After two-years working as a GIS technician, she enrolled the PhD program in Geography Department at the Florida State University in 2006.