Springer GIS / Cartography

Yuji Murayama *Editor*

Progress in Geospatial Analysis

Springer

Progress in Geospatial Analysis

Yuji Murayama
Editor

Progress in Geospatial Analysis

Springer

Editor
Yuji Murayama
Division of Spatial Information Science
Graduate School of Life and Environmental Sciences
University of Tsukuba, Tsukuba, Ibaraki, Japan

ISBN 978-4-431-53999-5 ISBN 978-4-431-54000-7 (eBook)
DOI 10.1007/978-4-431-54000-7
Springer Tokyo Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012941297

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book discusses the current trends in methods and applications of geospatial analysis and highlights future development prospects. It aims to provide a comprehensive discussion on data processing techniques, current practices with theories and models of remote sensing (RS) and geographical information systems (GIS), and empirical studies of geospatial analysis.

Data acquisition and processing techniques such as remote sensing image selections, classifications, accuracy assessments, models of GIS data, and the spatial modeling process are focused on in the first part of the book. In the second part, theories and methods, including fuzzy sets, spatial weights and prominence, geographically weighted regression, weight of evidence, Markov–cellular automata, artificial neural networks, agent-based simulation, multi-criteria evaluation, analytic hierarchy process, and a GIS network model are included. Part III presents selected best practices on geospatial analysis, focusing on geographical phenomena. Most of the chapters are original and a few, especially for applications, are reprints from international journals and proceedings.

This book is written for academicians, researchers, practitioners, and advanced graduate students. It is designed to be read by those new or starting out in the field of geospatial analysis, as well as by those who are already familiar with the field. The chapters are selected from experienced authors in such a way that readers who are new to the field will gain an important overview and insight. At the same time, those readers who are already practitioners in the field will gain from the advanced and updated materials and state-of-the-art developments in geospatial analysis.

Most of the contributors to this book are current faculty members, staff members, graduates, and PhD candidates of the Division of Spatial Information Science, University of Tsukuba, Japan. The Division, which was established in 2000 for including Geographical Information Science within the Doctoral Program in Geoenvironmental Sciences, provides an enabling research environment where faculty members, staff, and students work together to advance knowledge in GIS and remote sensing techniques in different areas of interest.

My sincere thanks go to the staff members of the Division of Spatial Information Science, University of Tsukuba, especially to Mr. Konstantin Greger, whose sharp eyes and skills with computers helped make the manuscript become a better book.

Tsukuba, Ibaraki, Japan Yuji Murayama

About the Editor

Yuji Murayama, Professor, Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan. Email: mura@geoenv.tsukuba.ac.jp, Web: http://giswin.geo.tsukuba.ac.jp/sis/en/faculty.html. Books published: Murayama, Y. and Thapa, R. B. eds. (2011). *Spatial Analysis and Modeling in Geographical Transformation Process: GIS-based Application*. Dordrecht: Springer, 300p. Kamusoko, C., Mundia, C. N. and Murayama, Y. eds. (2011). *Recent Advances in Remote Sensing and GIS in Sub-Saharan Africa.* New York: Nova Publishers, 211p. Murayama, Y. and Du, G. eds. (2005). *Cities in Global Perspective: Diversity and Transition*. Tokyo: College of Tourism, Rikkyo University with IGU Commission, 626p. Murayama, Y. (2000). *Japanese Urban System*. Dordrecht: Kluwer Publishers, 271p. Murayama, Y. (1991). *Spatial Structure of Traffic Flows.* Tokyo: Kokon-Shoin, 311p.

Contents

Contributors

Bingliang Cui School of Geography, South China Normal University, Guangzhou, People's Republic of China

Ronald C. Estoque Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Don Mariano Marcos Memorial State University, Bacnotan, La Union, Philippines

Courage Kamusoko Overseas Operations Division, Asia Air Survey Co. Ltd., Tokyo, Japan

Formerly in Integrative Environmental Sciences Program, University of Tsukuba, Tsukuba, Japan

Duong Dang Khoi Faculty of Land Administration, Hanoi University of Natural Resources and Environment, Vietnam

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Ko Ko Lwin Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Kondwani Godwin Munthali Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Yuji Murayama Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Rajesh Bahadur Thapa Earth Observation Research Center, Space Applications Mission Directorate, Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Changping Zhang Faculty of Regional Development Studies, Toyo University, Tokyo, Japan

Formerly in Graduate School of Geoscience, University of Tsukuba, Tsukuba, Ibaraki, Japan

Yaolong Zhao School of Geography, South China Normal University, Guangzhou, People's Republic of China

Formerly in Division of Spatial Information Science, University of Tsukuba, Tsukuba, Ibaraki, Japan

Chapter 1
Introduction: Geospatial Analysis

Yuji Murayama

1.1 What Is Geospatial Analysis?

Although the term "geospatial analysis" is widely used in the academic world, its definition is not clear even in the field of geographical information science (GIS). In this book, we define geospatial analysis as a GIS-based approach to analyze geographically referenced information using methods such as statistics, information theories, computational geometry, and geovisualization techniques. Its goal is to find the driving forces of changes on the earth's surface, and analyze the geographical phenomena in order to understand their processes and mechanisms. Finally, it allows to suggest appropriate policy planning and decision making for sustainable development of a region to be suggested.

"Geospatial" information is considered a part of "spatial" information, which receives wide recognition in GIS communities. Here, geospatial data are limited to those on the earth's surface which are geographically referenced by address, place name, latitude/longitude, and so on. Geological phenomena inside the earth, the free atmosphere far above the earth's surface, and the topography on the Moon and Mars, for example, are not encompassed by the category of geospatial data in this book. Furthermore, architectural concepts like building structure and room arrangement in houses are also outside our research territory, although they are embraced in spatial data. In this regard, field-based empirical studies using geospatial techniques are conducted mainly within the geography-related disciplines.

Y. Murayama (✉)
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: mura@geoenv.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_1, © Springer Japan 2012

1.2 Quantitative Revolution

The origin of geospatial analysis can be traced back to the 1950s, when in North America the contribution of academics to society in general was required by the penetration of practicalism (Murayama 2004a). Geography was criticized because its methodological framework was old-fashioned and insufficient to satisfy social needs. Against this backdrop, quantitative methods and theoretical models oriented to law-making were pursued, and these were applied in problem-solving studies (Murayama 2004b). This academic movement was called the *Quantitative Revolution* in geography (Burton 1963). New methodologies, including network approaches using graph theory, central place formulation, point pattern analysis, regionalization and classification techniques employing multivariate analysis, and location-allocation problems were developed (Berry and Marble 1968).

In the 1970s, geospatial analysis with a time dimension, that is, spatiotemporal analysis, emerged and quickly became popular. It embraces time geography, behavioral geography, studies of changes in land use/cover, spatial diffusion and interaction modeling, and so on. A strong emphasis was placed on the conceptual switch from static to dynamic thinking, in other words, from the spatial structure to spatial process studies. In the late 1970s, computer mapping was developed as a type of analytical cartography, paving the way for geovisualization studies linked with GIS (Tobler 1976).

1.3 GIS Revolution

In the 1980s and 1990s, new techniques in geospatial analysis were exploited and implemented in GIS software (Murayama 2001). Advanced geospatial analysis became available without programming via open-source freeware such as GRASS, SPATSTAT, and R Package. Methodologically, great interest was focused on spatial homogeneity and heterogeneity (Anselin 1988). This concept was developed by further studies on local spatial autocorrelation, geographically weighted regression (GWR), spatial weight metrics and prominence, Kernel density, modifiable area unit problems, trend surface, and so on.

Geospatial analysis is interdisciplinary in nature. To enhance its operability, strong cooperation among adjacent disciplines is indispensable. In the 1990s, sophisticated GIS software was invented through teamwork which brought together regional science, information science, computation geometry, and statistics. The analytical functions became powerful due to the integration of raster-based GIS and vector-based GIS. Furthermore, the combination of GIS with remote sensing (RS) enhanced the usability of geospatial analysis, where useful techniques in RS such as spatial metrics, land use/cover classification, fuzzy sets, and so on, were effectively introduced within the GIS framework (Murayama and Thapa 2011a).

Today, tremendous amounts of geospatial data are ceaselessly produced and have become immediately available in the GIS environment. This data-rich condition, along with increasing computing power, is enabling the development of advanced geospatial data mining. In the 1990s, geocomputation was developed with the use of fine image data from RS, and socioeconomic micro-data such as individual information from censuses and volunteered information (Openshaw and Abrahart 2000; Murayama and Thapa 2011a). The availability of high-resolution images from Quick Bird, IKONOS, and ALOS, for instance, is accelerating the micro-approach within cities, which is called urban remote sensing (Yang 2011). Today, geocomputational techniques are rapidly becoming strong tools for process-based studies where the future is predicted based upon the process from the past to the present. These techniques include neural networks, cellular automata, agent-based modeling, genetic algorithms, self-organizing mapping, and so forth. Since 1996, a geocomputation conference is held every 2 years, where geographers, information scientists, computer programmers, and other experts gather from all over the world to reflect on work within the field of geocomputation from an interdisciplinary perspective (http://www.geocomputation.org/).

The *GIS revolution* has brought methodological and conceptual shifts: from model-driven to data-driven, from deductive to inductive, from top down to bottom up, from aggregate to disaggregate, from discrete to continuous, from lagged time to real time, from static to dynamic, from quantitative to qualitative, and from linear to non-linear (Murayama and Thapa 2011b). From this revolution, it could be argued that geospatial analysis with GIS is more a useful tool in formulating hypotheses than it is in verifying them.

The book consists of 17 chapters. Chapter 1, the introduction, traces the evolution/origin of geospatial analysis using GIS-based techniques, and discusses the usefulness of theories, techniques, and methods for analyzing geographical processes occurring on the earth's surface. Then an overview of each chapter is given.

1.4 Geospatial Data Acquisition and Processing

In Part I, geospatial data acquisition and processing are discussed (Chaps. 2 and 3).

Chapter 2 focuses on the preprocessing of remotely sensed data and classification methods. Information extraction from satellite image data requires appropriate image-processing methods and techniques. The overall aim is to extract the information and explore the spatial patterns of land-surface objects, that is, land cover patterns from the classified satellite imagery known as a thematic map. Furthermore, this chapter presents theoretical procedures and techniques to produce useful information from such thematic maps. Significant technological advancements in data acquisition and analysis in recent years have made it easier to analyze the spatiotemporal dynamics of landscape changes using multispectral classification techniques with various statistical rules. Finally, this chapter discusses the various methods of assessing the classified image ranging from field-based validation to visual interpretation with high-resolution satellite images.

Chapter 3 focuses on geospatial data collection methods, database design and construction, and modeling with GIS. Geospatial data collection, including remote sensing, field surveying, and other in-house GIS data conversion processes (i.e., scanning, georeferencing, digitizing, etc.), is an important task for many geospatial information users. Traditional field data collection (i.e., pen-and-paper-based) is a bulky and time-consuming task. However, recent developments in mobile communication, global navigation systems, the Internet, and portable computational devices such as smartphone, netbooks or ultra-mobile personal computers (UMPCs) allow us to carry out field data collection in a timely manner. Moreover, under the client-server setting for field data collection, a field user may take advantage of digital repositories prepared for data collection (i.e., base maps, satellite images, and other ancillary data) as well as information resources more generally available via the Web. Proper geospatial data collection and conversion are required to support spatial analysis with GIS, which is vital for accurate decision making. Geospatial data processing is at the heart of the task in many GIS analyzes.

1.5 Geospatial Theories and Methods

GIS is designed to store, retrieve, manipulate, analyze, and visualize geospatial data. On the other hand, the uncertainty which affects the accuracy of maps and geospatial analysis results always exists in the data because of the limitations of human cognition of geographical phenomena or the resolving power of surveying instruments. Part II is composed of nine chapters (Chaps. 4 through 12), and aims to review the existing theories and methods in geospatial analysis, particularly in land use/cover modeling and GIS network data models.

Chapter 4 analyzes the fuzziness of geographical phenomena and classifies them into three aspects: the fuzziness of the distribution of the geographical phenomena or concept of a geographical entity, the fuzziness derived from spatial relationships, and the fuzziness derived from geospatial analysis and spatial reasoning operations. Fundamental fuzzy set theory, including the definition of a fuzzy set, fuzzy set operations, and fuzzy relationships and membership functions, is summarized in the representation of such fuzzy phenomena. Applications of fuzzy set theory in GIS are reviewed according to the sequence of fuzzy representations of geographical entities and their distribution, fuzzy spatial relationships, and fuzzy operations in spatial reasoning. Then, an example of a field-based integrated spatial reasoning model in the case of a constraint satisfaction problem (CSP) is given to synthetically illustrate the above-mentioned three fuzzy aspects. Finally, the author considers the future research directions of fuzzy set theory in applications using GIS.

Chapter 5 shows that in geospatial analysis of geographical phenomena, a region or a city under study might be divided into several small areal units such as a regular square tessellation or administrative units emerging in irregular shapes. The spatial interrelation between areal units can be expressed as different definitions of the weight coefficient. For example, the spatial structure of areal units might be defined as the spatial contiguity, which is treated as a spatial weights matrix ***W*** with a binary variable. This chapter also explains how to define and create the spatial weight, which is an expression of the spatial dependence between areal units. Four types of weight functions are introduced here. An area which has special geometric attributes and keeps significant spatial correlation close to adjacent areas is called a *prominent area*. As an application of the spatial weights matrix, the prominence of irregular areas can be measured by the *prominence index*, which is a stationary distribution of a Markov chain transition matrix that is identical to a spatial weight matrix. An empirical study in this chapter shows that generalized weight matrices are more appropriate for measuring the prominence rather than the distance decay and *k*-order.

Chapter 6 explains GWR, which is a technique for spatial statistical modeling used to analyze spatially varying relationships between geographical variables. Unlike the traditional regression framework, GWR allows local rather than global parameters to be estimated. In this chapter, the authors discuss the theoretical basis of the GWR method and modeling. Some of the current best practices for understanding urban and regional problems, for instance, regional analysis of wealth and land use/cover, driving forces behind deforestation and afforestation, and so on, are highlighted. GWR is one of the recent developments of local spatial analytical techniques, and it has been part of the growing trend in GIS toward local analysis.

Chapter 7, which deals with Bayesian theory, focuses on the weight of evidence (WofE) method and its applications in GIS. WofE, which is entirely based on the Bayesian approach of conditional probability, is traditionally used by geologists to point out areas which are favorable for geological phenomena such as seismicity and mineralization. Recently, the WofE method has been used to combine spatial data from a variety of sources to describe and analyze interactions, provide evidence for decision making, and construct predictive models. This chapter discusses the theoretical basis of WofE and presents best modeling practices. Basically, this method concerns the likelihood of detecting a certain event, which could be a given category of land use/cover change such as a change from an agricultural area to a built surface, in relation to potential evidence (proximity to urban centers, roads, water, etc.), often called the driving factors of change.

Spatial simulation models such as the Markov cellular automata (MCA) are critical for land use/cover change modeling because models are needed to gain insights into land use/cover change processes at many spatial and temporal scales. The MCA model combines cellular automata with Markov chains and GIS-based techniques such as multicriteria evaluation (MCE) and WofE in order to simulate land use/cover changes. The Markov chain process controls temporal dynamics among the land use/cover classes, while spatial dynamics are controlled by local rules determined either by the cellular automata mechanism or by its association with transition potential maps computed by WofE and MCE techniques. Chapter 8 reviews the methodological developments of the MCA model, as well its current status and future prospects based on criteria such as modeling techniques, data requirements, calibration, and validation. Thus, issues raised in this chapter could contribute to the improvement of future MCA land use/cover change models.

Chapter 9 outlines the progress of artificial neural networks, architectures, algorithms, and future developments in geospatial analysis with GIS. Their applications are reviewed with land use/cover change analysis and modeling. Many artificial neural network architectures have been developed over the past years. One of the most popular is the multilayer perceptron (MLP) neural network. From a geospatial analysis point of view, MLPs have been shown to be a universal and highly flexible function approximation tool for any data. This chapter gives a comprehensive review of the history and basic architecture of MLPs. The use of MLPs for land use/cover classification is presented as a representative type of data-pattern recognition. Finally, future trends in the development of MLPs are briefly summarized.

Chapter 10 reviews multiagent simulation models to understand land-use change management and sustainability in the area of forest loss (deforestation). An attempt is made to assess the sustainability of deforestation management from the perspective of individual decisions made at the farm or household level. Agent-based models (ABMs) continue to receive wide attention as a method of modeling complex real-world applications. Founded on the pretext of understanding the non-linearity of natural systems, the multidisciplinary ABM originates primarily from artificial intelligence. This chapter delves into the history of complex systems and connects it to present-day fundamental principles of defining and designing agents. It then highlights the basic tenets of agent modeling implementation, which include a discussion on available ABM development platforms. This chapter also contextualizes the agent, its environment, and its interaction when applications of ABM in land use/cover modeling are presented and addressed. All this adds to the available literature on ABMs in the hope that by presenting both the past and present, modelers will be able to distinguish between and develop new ABM approaches for the future.

Chapter 11 discusses multicriteria decision analysis (MCDA). Since its development in the 1970s, the analytic hierarchy process (AHP) has been an important tool for decision makers and researchers. It presents a flexible, step-by-step, and transparent way of analyzing complex problems in a MCDA environment based on experts' preferences, knowledge, and judgments. There has been a growing interest in using this method in the last two decades or so, and its scope of application has been expanding, especially in the field of geospatial analysis. This chapter also reviews the basic principles of AHP, its historical development, and its applications as a decision support tool for GIS-based MCDA. Major findings show that AHP has been implemented in various fields of geospatial analysis in various countries around the world. This indicates how versatile and useful AHP is as a decision support tool. However, the review also shows that researchers have still not achieved consensus on certain issues concerning the implementation of AHP as a weighting method for GIS-based MCDA. Some of these issues include considerations about the method of capturing expert opinion using the pair-wise comparison method, the method of aggregation of individual expert's ratings (in cases where consensus ratings are not used), and the method of standardizing the individual factors involved in a GIS-based MCDA. These issues are crucial; thus, careful attention is needed when using AHP as a decision support tool for GIS-based MCDA. Nevertheless, because of AHP's effectiveness in evaluating problems involving multiple and diverse criteria and the measurement of trade-offs, its simplicity and robustness, and its precision and ease of use, its applications will undoubtedly continue to expand in fields of both non-spatially and spatially based decision-making environments.

Chapter 12 discusses the GIS network model and its applications. A network is an interconnected set of points and lines that represent possible routes from one location to another. Road network models play a critical role in urban planning, emergency preparedness, retail market and market competition analysis, public facility management, and other planning and decision-making processes. Understanding the road network patterns in urban areas is important for human mobility studies, because people live and move along the road networks. Network data models allow us to solve daily activities such as finding the shortest path between two locations, looking for the closest facilities within a specific distance, and estimating driving time. A network model can include a multilayer model representing, for example, a railway system, a subway system, and a bus system to solve problems using multiple modes of transportation in an urban area. Many commercial GIS data models composed of layers such as points (nodes) and lines (links) comprise separate layers. This is called a layer-based approach. These nodes and links can also be represented as object classes; this type of model is known as an object-oriented network data model, and is still in the design phase. The development of 3D network models and concepts in GIS can solve these complex multilayer network solutions. Moreover, the combination of Internet technology and user-friendly Web-GIS provides an opportunity to perform interactive network analysis to make spatial decisions in a timely manner for local residents and city planners through Web-based GIS systems.

1.6 Applications in Geospatial Analysis

Part III provides five applications (Chaps. 13 through 17) in geospatial analysis employing GIS, remote sensing, and global positioning systems.

The complexity of urban systems requires integrated tools and techniques to understand the spatial process of urban development and project future scenarios. In this connection, Chap. 13 aims to simulate urban growth patterns using a Bayesian probability function in the Kathmandu metropolitan region in Nepal. Like many cities in the developing world, it has been facing rapid population growth and daunting environmental problems. Three time-series land-use maps at a fine scale (30 m resolution) derived from satellite remote sensing covering the last three decades of the twentieth century were used to clarify the spatial process of urbanization. Based on historical experiences of land-use transitions, the authors adopted the WofE method integrated in a cellular automata framework to predict the future spatial patterns of urban growth. The authors extrapolated urban development patterns to 2010 and 2020 under the current scenario across the metropolitan region. Depending on local characteristics and land-use transition rates, this model produced a noticeable spatial pattern of changes in the region. Based on the extrapolated spatial patterns, urban development in the Kathmandu valley is projected to continue through both infilling in existing urban areas and outward rapid expansion toward the east and south. Overall development will be greatly affected by the existing urban space, transportation networks, and topographic complexity.

Chapter 14 discusses land suitability assessment using a fuzzy MCE model, which is an important step for sustainable land-use planning in the rural landscapes of developing countries. Taking the Tam Dao National Park Region, Vietnam, as a case study, this chapter demonstrates how the results of the assessment are used for sustainable land-use planning decisions. Land degradation is recognized as one of the major threats to the buffer zones of protected areas (PAs) in Vietnam. In particular, the expansion of land degradation into the PAs is exerting pressure on biodiversity conservation efforts. This degradation is partially the result of mismanagement: the utilization of the land is often mismatched with the inherent nature of the land. Identification of the spatial distribution of suitable areas for cropland is essential to allow sustainable land-use recommendations to be made. The authors delineate the areas in the region which are suitable for cropland using a GIS-based MCE of biophysical factors and Landsat ETM+imagery. GIS is used to generate the factors, while MCE is used to aggregate them into a land suitability index. The results which indicate the location and extent of crop farming areas at different suitability levels can be used to identify priority areas for crop farming and sustainable land-use management. The GIS–MCE approach provides an effective assessment tool for land-use managers working in the PAs of Vietnam.

Chapter 15 presents the application of the neighborhood interaction method in land-use analysis. Local spatial interactions between neighborhood land-use categories (i.e., neighborhood interactions) are important factors which affect urban land-use change patterns. Therefore, they are key components in cellular automata-based urban geosimulation models that aim at forecasting urban land-use changes. In this chapter, the authors interpret the similarities and differences of neighborhood interactions in three metropolitan areas in Japan, namely Tokyo, Osaka, and Nagoya, to provide empirical material to understand the mechanism of urban land-use changes. Neighborhood interaction reveals the effect of spatial autocorrelation in the spatial process of urban land-use changes in the three metropolitan areas, which correspond with the agglomeration of urban land-use allocation in Japan. Neighborhood interactions amidst urban land-use changes among the three metropolitan areas generally showed similar characteristics. The regressed neighborhood interaction coefficients in the models may represent the general characteristics of the neighborhood effect on urban land-use changes in cities. The results provide very significant material for exploring the mechanism of urban land-use changes, and the construction of universal urban geosimulation models which may be applied to any city.

Walkability is a well-known measure of how conducive an area is to walking to and from chosen destinations. The calculation of a walk score is widely used in accessibility studies to determine the ease or difficulty of travel by foot between one point and another. Based on this situation, Chap. 16 proposes an integrated methodology (remote sensing, GIS, and spatial Web technology) to model urban green space walkability which will enable local residents to make informed decisions that will improve their living conditions and physical health in relation to the neighborhood environmental quality. The authors discuss the modeling of urban green space walkability by utilizing a Web-based GIS to calculate eco-friendly walk scores based on the presence of green spaces by integrating advanced land observing satellite data and other GIS datasets. We use this spatial Web technology to help local residents make decisions related to neighborhood environmental quality, such as how to choose an eco-friendly living space when buying a home, or how to find the shortest or greenest route to walk to improve their health.

Chapter 17 deals with watershed evaluation using geospatial techniques. Severe watershed degradation continues to occur in the tropical regions of southern Africa. This has raised interest in harnessing and manipulating the potential of watershed resources for human benefit as populations grow. The Songwe river is one such degrading watershed that causes biennial flooding, among other problems. In this study, climatic, land use, topographic, and physiographic properties were assembled for this watershed, and were used in a process-based GIS with the aim of determining the hydrological sediment potential of the Songwe river watershed and quantifying the possibilities of reservoir sedimentation. The study further aimed at determining the critical sediment-generating areas for prioritized conservation management, and the relationship between the increasing flood events in the floodplains and the rainfall trends. Based on the evaluation of hydrological runoff processes using the Pan-European Soil Erosion Risk Assessment (PESERA) model, the estimated amount of sediment transported downstream is potentially huge. It was established that most of the sediment generation was occurring in the upper sub-basin, and specifically from built-up villages and degraded natural land. These trends have not only caused an increase in flooding events in the lower sub-basin, but also pose a great sustainability risk of sedimentation to the proposed reservoir.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic, Dordrecht

Berry BJL, Marble DF (1968) Spatial analysis: a reader in statistical geography. Prentice-Hall, Englewood Cliffs

Burton I (1963) The quantitative revolution and theoretical geography. Canadian Geographer 7 (4): 151–162

Murayama Y (2001) Geography with GIS. GeoJournal 52:165–171

Murayama Y (2004a) How has GIS contributed to urban geography? Annu Rep 30:5–8, Inst. Geosci., Univ. Tsukuba

Murayama Y (2004b) American influence on Japanese human geography: a focus on the quantitative and GIS revolutions. GeoJournal 59:73–76

Murayama Y, Thapa RB (2011a) Spatial analysis and modeling in geographical transformation process: GIS-based applications, vol 100, GeoJournal library series. Springer, Dordrecht

Murayama Y, Thapa RB (2011b) Spatial analysis: evolution, methods, and applications. In: Spatial analysis and modeling in geographical transformation process: GIS-based applications, vol 100, GeoJournal library series. Springer, Dordrecht, pp 1–26

Openshaw S, Abrahart RJ (2000) GeoComputation. CRC Press, London

Tobler W (1976) Analytical cartography. American Cartographer 3:21–31

Yang X (2011) Urban remote sensing: Monitoring, synthesis and modeling in the urban environment. Wiley, Chichester

Part I
Geospatial Data Acquisition and Processing

Chapter 2
Multispectral Classification of Remote Sensing Data for Geospatial Analysis

Duong Dang Khoi and Kondwani Godwin Munthali

2.1 Introduction

Remote sensing data are one of the primary data sources for many geospatial analyzes. The nature of remote sensing data acquisition ranges from ground-based to airborne to space-borne. There are two types of remote sensing: active and passive. Passive remote sensing sensors detect the natural radiation that is emitted from, or reflected by, the object or surrounding area being observed. Reflected sunlight is the most common source of radiation measured by passive sensors. Some examples of passive remote sensing satellites are Landsat MSS/TM/ETM+, SPOT, IKONOS, QuickBird, etc. Active remote sensing emits energy in order to scan objects, and then detects and measures the radiation that is reflected or back-scattered from the target. Radio detection and ranging (RADAR), light detection and ranging (LiDAR) and sound navigation and ranging (SONAR) are examples of active remote sensing where the time delay between emission and return is measured, thus establishing the location, height, speed and direction of an object.

In the 1860s, the observation of the earth using a balloon became the starting point of what would later be called remote sensing (Lillesand et al. 2008). Observation of the earth from airborne platforms has a history of 150 years, but most of the technical

D.D. Khoi (✉)
Faculty of Land Administration, Hanoi University of Natural Resources and Environment, Vietnam

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: khoi_tn@yahoo.com

K.G. Munthali
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kmunthali@gmail.com

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_2, © Springer Japan 2012

innovation and development has taken place in the last three decades. The launch of the first remote sensing satellite by the United States in 1972 paved the way for applications of remote sensing in studies of the earth's resource management, including the management of forests. Satellite imagery is increasingly used in various fields such as agriculture, forestry, geology, hydrology, land use/cover change, oceans and coastal monitoring. Remote sensing images are processed and combined with other ancillary datasets in geographical information science (GIS) via spatial analysis techniques to provide significant information for environmental monitoring and management.

This chapter presents a multispectral classification process for space-borne remote sensing data, which is the most important part of information extraction from satellite images. Here, we explain the complete procedures of multispectral classification with image preprocessing, image enhancement, supervised and unsupervised classification, training-site selection, a setup classification algorithm and methods of accuracy assessment.

2.2 Multispectral Classification Procedures

Landsat satellite images contain information in digital numbers, and therefore a classification procedure is required to transform those digital numbers into understandable geographic features. This is known as information extraction. An image processing procedure can be defined as a process of extracting distinct geographic features or categories from satellite images based on supervised or unsupervised classification methods. The unsupervised method is the division of the whole image into different categories based on the similarity of spectral signatures, where each category is labelled with a specific name. In contrast, the supervised classification method uses prior knowledge (such as existing land-use maps, ground-based observations and aerial photographs) to classify the images into geographic feature patterns. This is called thematic mapping.

2.2.1 Preprocessing

Raw digital images usually contain geometric distortions so they cannot be used directly as a map base without subsequent processing. Preprocessing is an important and diverse set of image preparation steps that act to offset problems with satellite image band raw data, and recalculate the values of digital numbers to minimize problems with that data. It is difficult to decide what should be mentioned under the topic of preprocessing, since the meaning of what is or is not a deficiency in the data depends on the extent or scope of the application. Generally, preprocessing comprises a series of sequential operations, including atmospheric correction, image registration, geometric correction and masking.

The first category is the adjustment of radiance error represented by digital numbers (DN) and caused by factors such as noises (band stripping, scan line drop-out and salt-and-pepper error), and atmospheric and topographic effects. *Stripping* means errors that occur in the sensor response and/or data recording and transmission which result in a systematic error or shift of pixels between rows. This occurs when a sensor goes out of adjustment and produces readings that are higher or lower than other sensors at the same band. The procedure for correcting stripping is known as *de-stripping*. This involves the calculation of the mean and standard deviation for the entire image. The output from each sensor is scaled to match the mean and standard deviation of the entire image, and therefore the value of each noise pixel in the image is corrected or adjusted. *Scan line drop-outs* are errors that occur in the sensor response and/or data recording and transmission which mean the loss of a row of pixels in the image. This issue can be solved in several different ways. Reclassification of the affected bands and the use of filter operations are good examples. In addition, a *salt-and-pepper error* is a random noise which causes the values to be high or low relative to surrounding pixel values. Median filter running can be used to remove this error.

The atmosphere affects satellite imagery in different ways. Principally, solar radiance when traveling to the surface of the earth is scattered and absorbed by various gaseous substances such as carbon dioxide, oxygen, ozone and haze. Various procedures and models are available in commercial GIS/RS software systems, such as IDRISI Taiga, for removing these atmospheric effects. Topographic effects are also important. The interaction of the angle and the azimuth of the sun's rays with slope and aspect produce such effects. For instance, if an image is recorded in the early morning, the effect of the sun's angle on slope illumination is extreme. Or in mountain areas, reflectance from sloping areas facing the sun's rays is often lower than the overall reflectance of the entire area. Several techniques can be employed for correcting topographic effects, such as band ratio, image partitioning and illumination modeling. These procedures are easily implemented in IDRISI software.

The second category relates to geometric restoration of the imagery according to a given reference map, such as a published topographic map or GPS-based ground control points. *Resampling* is commonly used to produce better estimates of the values of digital numbers for individual pixels adjusted to real spatial locations. The accurate registration of satellite imagery is essential for analyzing remote sensing data from a particular area. Each image should be georeferenced to the topographic map of the study area, and is resampled using the nearest-neighbor algorithm.

Image enhancement techniques can improve the quality of an image as perceived by a human. These techniques are very useful, because many satellite images when examined on a color display give inadequate information for image interpretation. There is no conscious effort to improve the fidelity of the image with regard to some ideal form of that image. However, a wide variety of techniques exist for improving image quality. Contrast stretch, density slicing, edge enhancement and spatial filtering are the most commonly used techniques. Image enhancement is attempted after the image has been corrected for geometric and radiometric distortions. Image enhancement methods are applied separately to each band of a multispectral image. Digital techniques have been found to be more satisfactory than photographic techniques for image enhancement because of the precision and wide variety of digital processes.

The visualization of spatial patterns of digital numbers in the images is very important for understanding major features or patterns dominant within the entire region. Individual bands are carefully visualized to improve the interpretability of the image. Band visualization provides an understanding of the spectral patterns of an image. Each of the Landsat imagery bands can provide unique information for the interpretation of surface features. For example, Band 1 provides information about the penetration of water bodies, and thus is able to differentiate soil and rock from vegetation and detect cultural features. Band 2 is sensitive to differences in water turbidity. This band can separate vegetation types, e.g. forest and cropland, from soil. In this band, settlements and infrastructures have a brighter tone, while vegetation has a darker tone. Band 3 is a spectral region of strong chlorophyll absorption, and therefore this band can distinguish between vegetation and soil. It is capable of separating primary forest, secondary forest and cropland areas. Band 4 can distinguish vegetation and its conditions and is therefore able to separate primary from secondary forest (degraded forest). Water bodies are a strong absorber of near infrared energy, and therefore this band clearly delineates water bodies and separates dry and moist soils. Band 5 is capable of separating forest, cropland and water bodies, as forest has a darker tone than cropland, while water bodies have an even darker tone than forest or cropland. Band 7 has the capacity to separate secondary forest from primary forest areas. Aside from the visualization of individual bands, composite images are also employed to enhance the interpretability of features of the images.

2.2.2 Unsupervised Classification

Sometimes, observations or secondary data of land-cover types are not available for a particular study area. Even the number of categories can be unknown. In such situations, it is impossible to estimate the spectral reflectance means of the categories. Therefore, unsupervised classification, known as automatic classification procedure, should be used.

Unsupervised classification aims at separating the raw image data into a more human-readable form using as little operator input as possible. This automated classification method creates a thematic raster layer from a remotely sensed image by letting the software identify statistical patterns in the data without a priori knowledge of the area (Leica Geosystems Geospatial Imaging 2005; Lillesand et al. 2008). It employs the iterative self-organizing data analysis technique algorithm (ISODATA) to arbitrarily cluster the means of an existing signature set, moving the means each time the algorithm iterates. The iterative procedure terminates when either the number of iterations set when running the algorithm has been reached, or a convergence threshold—the maximum percentage of pixels whose class allocations are allowed

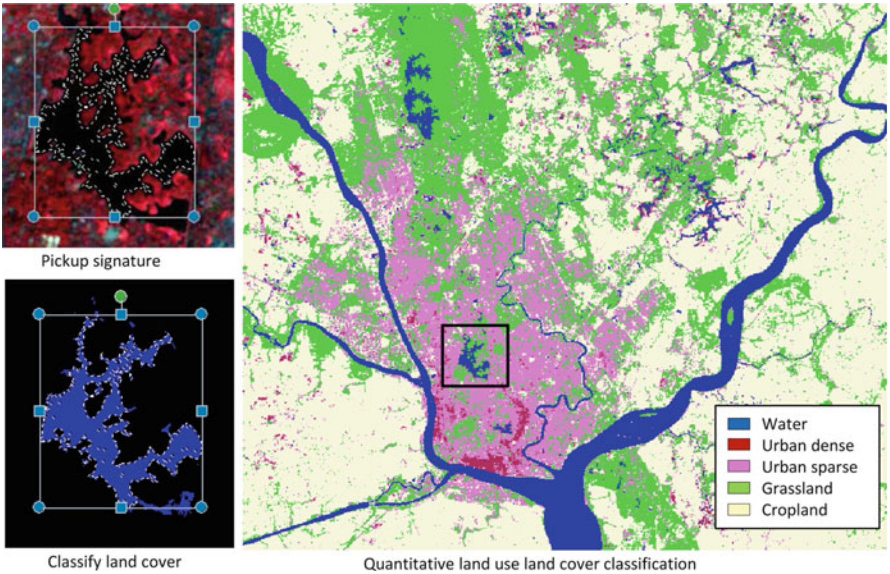


Fig. 2.1 Example of signature development for multispectral supervised classification of Landsat ETM, Yangon City (Myanmar) (*Source*: Lwin et al. 2012)

to remain unchanged between iterations—is reached (Leica Geosystems Geospatial Imaging 2005). Traditionally, the number of output clusters from an unsupervised classification is larger than the actual number of classes required for that application. A posteriori knowledge is then used to merge spectrally similar classes, and label them accordingly, to generate a thematic land use/cover map.

2.2.3 Supervised Classification

Unlike unsupervised classification, supervised classification is based on prior knowledge of the area shown in the image. The process of supervised classification may follow several steps (Fig. 2.1), which are briefly summarized into two stages. The first stage is the identification of categories of real-world features, i.e. land use/cover types, using the prior knowledge of features in a study area from both primary and secondary data. This step is known as the delineation or identification of training areas. The second stage is the labeling of classified categories using a selected classification rule.

Delineation of training areas: Training areas (TA) are known areas of a land use/cover category determined by ground observations or by an inspection of aerial photographs and the reference land use/cover maps that are assumed to be true information. In the image, TAs are represented as a set of pixels for the known area of that land use/cover class. The delineation of TA is most effective when the image interpreter has full knowledge of the study area and a good reference land use/cover map.

Signature evaluation: A spectral signature (signature) is a set of pixels (Fig. 2.1) that statistically defines a training site set for a specific land use/cover type. It is good practice to generate spectral signature files from the largest possible number of bands. Each signature is defined by statistical parameters, including the number of bands, the minimum, maximum, and mean values of the training areas, the covariance matrix of the training areas and the number of pixels in the training areas.

Classification rules: The effective classification of remote sensing data depends on the separation of land use/cover types into a set of spectral classes (signatures) that represent the data in a form which is suitable for a particular classifier algorithm.

Several mathematical algorithms can be used as supervised classification procedures or decision rules, and also to assign an unknown pixel to one of a number of classes. The choice of a particular classifier depends on the nature of the input data and the desired output. Parametric classification algorithms assume that observed measurement vectors for each class in each spectral band are normally distributed. Non-parametric classification algorithms, on the other hand, make no such assumption. Several classification algorithms can be employed, such as parallelepiped, minimum distance and maximum likelihood.

The parallelepiped classification algorithm is a widely used decision rule based on simple Boolean "and/or" logic. Training data in spectral bands are used when performing the classification. The brightness values from each pixel of the multispectral imagery are used to produce a multidimensional mean vector. The decision boundaries form a multidimensional parallelepiped in feature space. If the pixel value lies above the lower threshold and below the high threshold for all bands evaluated, it is assigned to an unclassified category.

In the second approach, the geographical principle that two nearby objects are likely to be similar is used in a nearest-neighbor simple classifier. Here, the classifier simply finds the closest object from the training set to an object being classified in the N-dimensional feature space (White 1996). The strongest advantage of nearest-neighbor algorithms is that they are easy to implement, and do provide satisfactory results if the features are chosen and weighted carefully. However, they do not simplify the distribution of the objects in the parameter space sufficiently clearly when the training set is retained in its entirety as a description of the object distribution. It is also very sensitive to irrelevant parameters, thereby compromising its results significantly (White 1996).

In the case of maximum likelihood, the classifier assumes a special probability distribution, for instance a Gaussian distribution, of the given data a priori, and then determines the appropriate parameters from the training data (Keuchel et al. 2003). Each data pixel is then assigned to the class for which its values are most likely, i.e. the class with the highest a posteriori probability (Swain and Davis 1978). The maximum likelihood algorithm is commonly employed in the separation of land use/cover classes. This method is useful because it requires a minimum of training area data while achieving high accuracy. The image interpreter trains the software to recognize spectral values associated with the training areas. After the signatures for each land use/cover have been defined, the software uses those signatures to classify the remaining pixels.

Other classification methods: In a case where the raw image data have a mixed pixel or heterogeneous feature representation so that an individual pixel cannot be definitively assigned to one category, a membership function is used, where a pixel's value is determined by whether it is closer to one class than another (Jensen 2005; Wang 1990). This is called fuzzy classification (or fuzzy supervised classification). Each classified pixel is then assigned a membership grade with respect to its membership in each information class. This generates two maps, a multilayer class and distance map. The fuzzy convolution utility then uses these two output maps to create a single classification layer by calculating the total weighted inverse distance of all the classes in an $n \times n$ window of pixels by assigning the center pixel in the class with the largest total inverse distance summed over the entire set of fuzzy classification layers (Leica Geosystems Geospatial Imaging 2005).

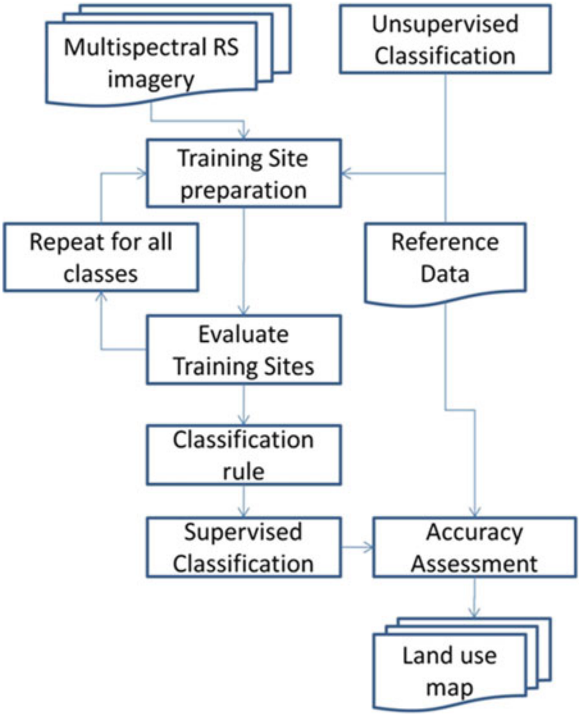
In the hierarchy of classification methods, neural networks are probably the most widely known. These are a family of artificial intelligence techniques whose design and development was inspired by biological neural networks (Knight 1990; Lippmann 1987). Both perceptron and back-propagation neural network algorithms start from random initial states, and use a training set containing a representative sample of patterns of each class to "teach" the network to perform the desired classification function (Odewahn et al. 1992). The perceptron is a simpler classifier that forms a hyperplane parameter space to separate two classes by forcing the trained classification to converge to a vector solution if the classes are linearly separable (Duda and Hart 1973). The back-propagation network is the more complex of the two and is capable of learning much more complex functions without restricting itself to linear separability problems (Odewahn et al. 1992). Rumelhart and McClelland (1988) present a detailed account of the latter. The downside of the use of neural networks, apart from being computationally slow, is that it is difficult to determine how the network makes its decisions, and consequently it becomes difficult to determine which parameters are important to the classification, apart from being computationally slow (White 1996).

A hybrid classification method is a combination of any of these classification methods. It has been employed in cases where any one of the classification approaches alone would not produce good results. Figure 2.2 shows an example of a hybrid multispectral image classification.

2.2.4 Post-classification Processing

Post-classification processing is very important for removing any mismatched locations of classified land use/cover. The slope and elevation data, for instance, have been used to examine the correctness of all land use/cover classes in land use/cover

Fig. 2.2 Multispectral hybrid classification workflow combining supervised and unsupervised approaches



classification applications. For example, the pixels of paddy fields with a very steep slope can be considered as mismatched locations. These locations can be re-assigned into the forest class. This also helps to keep in check confusing scene selection during the classification where, for instance, crop land and bare land tend to have similar spectral reflectance during dry seasons, just as paddy fields and grassland do during the growing or rainy season.

2.2.5 Accuracy Assessment

The next critical step in the process of evaluating the quality of land use/cover map production is an assessment of accuracy. Accuracy assessment is a comparison between a classified map and a reference map that is assumed to be true (Lillesand et al. 2008). The most common way to express classification accuracy is the preparation of a so-called error matrix, which is also known as a confusion matrix or a contingency matrix. Such matrices show a cross-tabulation of the classified land use/cover and the actual land use/cover according to the results of sample site observations. The matrix lists the values for known land use/cover types of the reference data in the columns and those for classified land use/cover data in the rows. The main diagonal of the matrix shows the correctly classified pixels. The producer's accuracy, user's accuracy, overall accuracy and κ-statistic are calculated in order to evaluate the land use/cover classification performance.

The overall accuracy represents the percentage of correctly classified samples. This is calculated by dividing the correctly classified pixels (the sum of the values in the main diagonal) by the total number of pixels checked by reference maps/aerial photographs or observed in the field, and is given as follows:

$$\frac{D}{N} \times 100\%, \tag{2.1}$$

where D is the total number of correct pixels summed along the major diagonal, and N is total number of pixels (site observations) in the error matrix.

Aside from the overall accuracy, the classification accuracy of individual classes can be calculated in a similar way. The producer's accuracy and the user's accuracy are possible. The producer's accuracy is a measure of omission errors that correspond to those pixels belonging to the class of interest that the classifier has failed to recognize. It is derived by dividing the number of correct pixels in one class by the total number of pixels derived from the corresponding reference data class (the column total):

$$\frac{X_{ii}}{X_{+i}} \times 100\%, \tag{2.2}$$

where X_{ii} is the total number of correct cells in a land use/cover class, and X_{+i} is the sum of cell values in the column.

In a similar way, the correctly classified pixels in a class are divided by the total number of pixels that are classified in that class to get a measure called the user's accuracy. The user's accuracy is a measure of the commission errors (Richards and Jia 1999), or simply the measure of the reliability of the map. It informs users about how well the map represents what is really on the ground. This measure is calculated as follows:

$$\frac{X_{ij}}{X_{+j}} \times 100\%, \tag{2.3}$$

where X_{jj} is the total number of correct cells in a land use/cover class, and X_{+j} is the sum of the cell values in the row.

The κ-coefficient is a measure of the overall agreement of a matrix. In contrast to the overall accuracy, i.e. the ratio of the sum of the diagonal values to the total number of cell counts in the matrix, the κ-coefficient also takes into account non-diagonal values. The κ-coefficient has become a standard component of most accuracy assessments (Congalton 1991; Hudson and Ramm 1987). It is given as follows:

$$\hat{K} = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} X_{i+}X_{+i}}{N^2 - \sum_{i=1}^{r} X_{i+}X_{+i}}, \tag{2.4}$$

where r = the number of rows and columns in the error matrix, N = the total number of observations (pixels), X_{ii} = the observation in row i column i, X_{i+} = the marginal total of row i, and X_{+i} = the marginal total of column i.

The accuracy of the classifications is assessed using the same reference sample. Using a stratified sampling method, a set of reference pixels per class can be randomly selected for the accuracy assessments. Randomly selected reference pixels lessen or eliminate the possibility of bias (Congalton 1991). The sample size for each land use should be based on the spatial extent of each land use/cover type. A geographical position system (GPS) and a digital camera are often used to collect the site data and record views of the sites for analysis. Then an error matrix for the classification is generated by visually and carefully interpreting each sample pixel.

Other forms of accuracy assessment involve the use of high-resolution satellite images, for example by identifying the shapes of land use/cover categories such as grassland, roof patterns, vegetated areas, etc. in a classified image and visually comparing these side by side with the corresponding categories in a high-resolution image (Fig. 2.3). In this approach miss-classified pixels can be detected, for example when the shadows of roofs and buildings have been classified as water (blue areas) in Fig. 2.3.

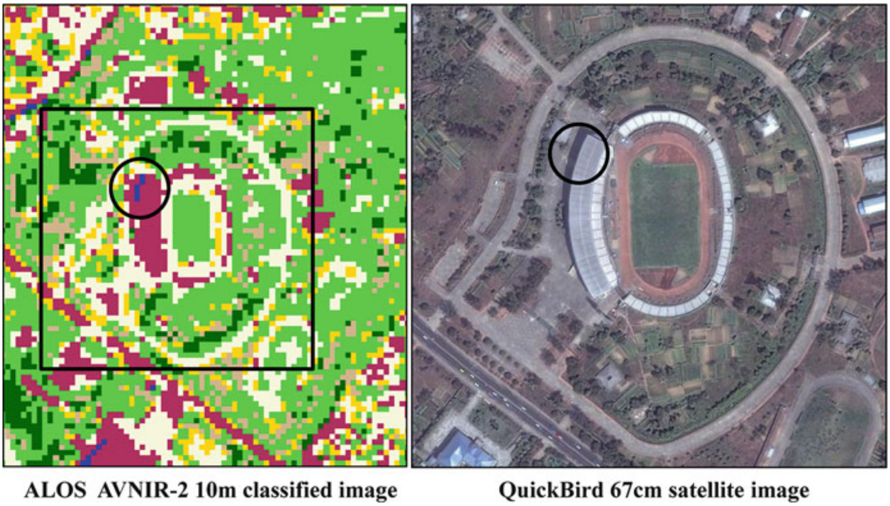


Fig. 2.3 Use of a high-resolution satellite image (right) for accuracy assessment of a classified image (left) (*Source*: Lwin and Murayama 2011)

For an analysis of changes in land use or land cover, aerial photographs using classified images from the previous year are important for accuracy assessments. This involves scanning, georeferencing and digitizing the photographs, as they are usually in the form of hard copies. Sometimes, asking the native people of the area helps to determine what land cover was there in the past.

2.3 Application in Geospatial Analysis

Multispectral classification of space-borne remote sensing data has a plethora of applications in geospatial analysis. Although classification is not usually an end in itself, it is evident that it is an important aspect, as has been shown in selected works relating to land use/cover changes, especially urbanization and deforestation.

In the hope of developing an understanding of urban growth in rapidly changing environments in sub-Saharan Africa, Mundia and Murayama (2011) embarked on a project to model urban growth in Kenya's capital, Nairobi. Of particular interest to them was the estimation of future outcomes of current spatial plans and policies for land-use development, and the consideration of alternative planning and policy scenarios in order to minimize their impact. To achieve this they needed accurate projections of future urban growth, which in turn needed a solid foundation of quality analysis and a good understanding of the likely patterns and trends of urban change (Abiodun 1997; Rakodi 1997; Hope and Lekorwe 1999). They used the Clarke cellular automata (CA) urban growth model (Clarke and Gaydos 1998), which they modified and calibrated accordingly, and also geographical information system (GIS) modules. The latter allowed GIS analyzes to be used to determine suitability factors, model constraints and land use/cover change, while the former was useful for model calibration and for applying transitional change rules (Mundia and Murayama 2011). To map the land use/cover changes they used post-classification techniques on multispectral Landsat images for 1976, 1988 and 2000. The classified maps obtained for these years are as shown in Fig. 2.4. Although the land use/cover maps were just part of a series of input data for the CA urban growth model, the importance of multispectral classification techniques cannot be overemphasized, especially in this case where problems with data availability and accuracy in Nairobi would have made the analysis difficult. With satellite data widely available and routinely collected, the urban growth analysis using CA was satisfactorily achieved in such a data-sparse environment by employing classification techniques (Mundia and Murayama 2011).

It has been reported that two-thirds of the population living in the rural areas of central and southern Africa depends on agriculture and other natural resources such as timber and firewood for their economic and social needs (Campbell et al. 2000; Gambiza et al. 2000). As a consequence, an understanding of land use/cover change dynamics is fundamental for rural land-use planning, especially if sustainable agriculture and forestry management are to be achieved. With this background Kamusoko et al. (2009) simulated future land use/cover changes (up to 2030) in the

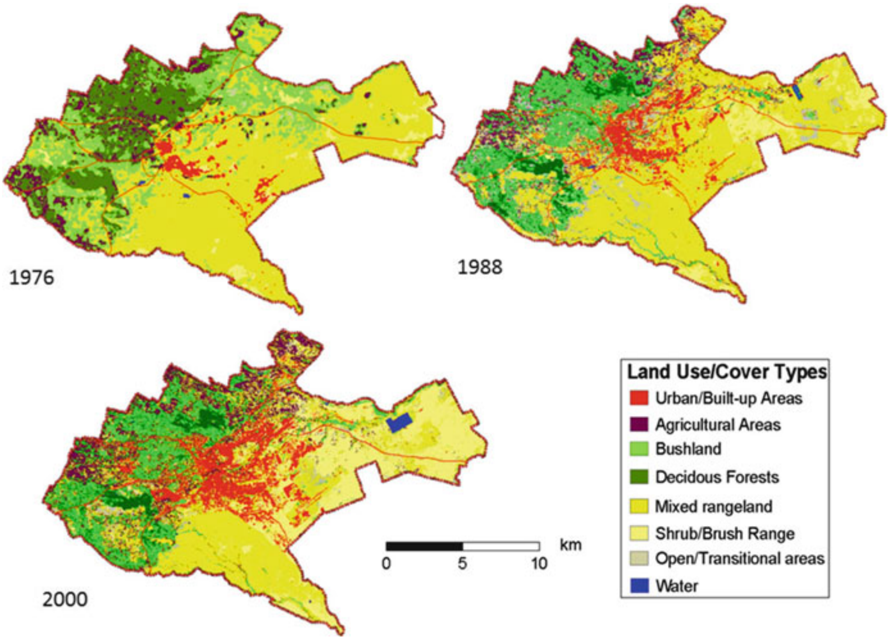


Fig. 2.4 Land use/cover maps of Nairobi for 1976, 1988 and 2000 (*Source*: Mundia and Murayama 2010)

Masembura and Musana communal areas of the Bindura district, Zimbabwe, based on the Markov–cellular automata model that combined Markov chain analysis and cellular automata models. Using the Markov chain analysis, they computed transition probabilities from the Landsat classification-derived land use/cover maps (1973, 1989 and 2000). The land use/cover maps for 1973, 1989 and 2000 were extracted from the Bindura district land use/cover maps that were classified from Landsat data based on a hybrid supervised/unsupervised classification approach coupled with GIS (Kamusoko and Aniya 2007). Together with other biophysical and socioeconomic data, the land use/cover maps derived were then used to (1) define initial conditions, (2) parameterize the Markov–cellular automata model, (3) calculate transition probabilities, and (4) determine the neighborhood rules with transition potential maps (Kamusoko et al. 2009). The results showed that if the current land use/cover trends in the study area continue without holistic sustainable development measures, severe land degradation will ensue.

In the bowl-shaped valley of Kathmandu, Nepal, rapid demographic and environmental changes and weak land-use planning practices in previous decades have resulted in environmental deterioration, haphazard landscape development, and stress on the ecosystem structure (HMGN/UNCTN 2005). As a consequence, a lot

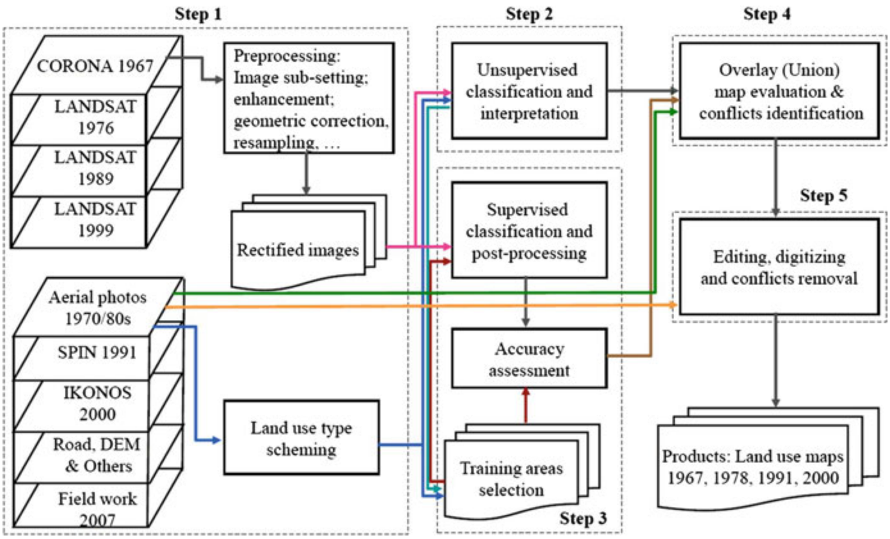


Fig. 2.5 Land-use mapping scenario for remote sensing images (*Source*: Thapa and Murayama 2009a)

of agricultural and forest land has been converted into urban areas and human settlements (Thapa et al. 2008). Therefore, with the aim of monitoring the urbanization and environmental consequences in the valley, Thapa and Murayama (2009a) examined the spatiotemporal pattern of urbanization in the valley using remote sensing and spatial metrics techniques by quantifying land-use patterns and analyzing the changes over time. Remote sensing not only provides spatially consistent data sets that cover large areas with both high spatial details and high temporal frequency (Jensen 2007; Thapa and Murayama 2009b), but it also provides consistent historical time series that date back to the 1960s. Although the satellite image data are detailed, many operational and applied remote-sensing analyzes still require the extraction of discrete thematic land-surface information from these satellite images using classification-based techniques (Mesev 2003; Yuan et al. 2009), as the satellite image data are not provided in an objective thematic format (Thapa and Murayama 2009a). It is argued that the heterogeneity and complexity of the landscape in urban regions require land use/cover classification techniques that combine more than one classification procedure in order to improve the accuracy of remote sensing-based mapping (Thapa and Murayama 2009b; Prenzel and Treitz 2005). Therefore Thapa and Murayama (2009a) followed a series of processing steps (Fig. 2.5) to transform the satellite data into meaningful thematic information. Their results showed that the urbanization process has resulted in fragmented and heterogeneous land-use combinations in the valley, with urban/built-up areas having grown slowly in the 1960s and 1970s, but having grown rapidly since the 1980s (see Thapa and Murayama 2009a).

In the protected area of the Tam Dao National Park in Vietnam, Khoi and Murayama (2010) demonstrated the application of remote-sensing data and a coupled neural network–Markov model for a predictive model of deforestation in the area. In this case study, remote-sensing data were used to interpret and understand spatial patterns of land use/cover changes, especially changes to forests in the area. Then the biophysical and socioeconomic drivers of forest change were investigated. Finally, a neural network–Markov model was employed to predict deforestation patterns in the area in the years 2007, 2014 and 2021. This approach provides an effective tool for monitoring deforestation at both small and large scales. In the context of protected areas in developing countries, prediction maps of changes in forest patterns can help the managers of protected areas to identify where conservation and forest management efforts should be focused.

2.4 Conclusions

While the monitoring of land use/cover changes is needed to understand and predict the dynamic process of land use/cover patterns at different times, the traditional labor-intensive survey/fieldwork approaches are limited in their effectiveness, as they are often unable to reveal the spatial patterns of landscape changes and the environmental consequences that occur in a given time period (Thapa and Murayama 2009a). However, significant technological advances in data acquisition and analysis techniques in recent years have made it easier to analyze the spatiotemporal dynamics of landscape changes (Herold et al. 2003). With these developments in techniques, remotely sensed images from airborne and satellite sensors provide a large amount of cost-effective, multispectral and multitemporal data to monitor landscape processes and estimate the biophysical characteristics of land surfaces (Miller and Small 2003; Herold et al. 2003; Jensen 2007; Thapa and Murayama 2008). The ability to handle and process remote-sensing data has resulted in the extraction of objective thematic mapping of the form, land uses and density of geographic features, each of which has an associated shape, configuration, structure, pattern and organization (Miller and Small 2003; Lillesand et al. 2008; Thapa and Murayama 2009b). Satellite imagery coupled with advanced handling and analysis techniques has the unique ability to provide synoptic views of large areas at a given time which are not possible using conventional survey methods (Thapa and Murayama 2009a). It is therefore not surprising that a wide range of remote sensing applications from both active and passive sensors is currently available. These include quantifying urban growth and land-use dynamics (Weber and Puissant 2003; Herold et al. 2003), landscape pattern analysis (Zhang 2005; Yu and Ng 2006; Li and Yeh 2004), urbanization (Weng 2007; Thapa and Murayama 2009a; Mundia and Murayama 2010), socioeconomic applications (Seto and Kaufmann 2003; Kamusoko et al. 2009), life quality improvement (Gatrell and Jensen 2008), urban infrastructure characterization (Mesev 2003), microclimate and hydrology (Carlson and Arthur 2000), and topographic mapping (Zandbergen 2008).

References

Abiodun JO (1997) The challenges of growth and development in metropolitan Lagos. In: Rakodi C (ed) The urban challenge in africa: growth and management of its large cities. United Nations University Press, Tokyo, pp 192–222

Campbell BM, Costanza R, van den Belt M (2000) Land use options in dry tropical woodland ecosystems in Zimbabwe: introduction, overview and synthesis. Ecol Econ 33:341–351

Carlson TN, Arthur TS (2000) The impact of land use – land cover changes due to urbanization on surface microclimate and hydrology: a satellite perspective. Global Planet Change 25:49–65

Clarke KC, Gaydos L (1998) Loose coupling a Cellular Automata model and GIS: long-term growth prediction for San Francisco and Washington/Baltimore. Int J Geogr Inf Sci 12:699–714

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37:35–46

Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley, New York

Gambiza J, Bond W, Frost PGH, Higgins S (2000) Special section: land use options in dry tropical woodland ecosystem in Zimbabwe. A simulation model of Miombo woodland dynamics under different management regimes. Ecol Econ 33:353–368

Gatrell JD, Jensen RR (2008) Sociospatial applications of remote sensing in urban environments. Geography Compass 2:728–743

Herold M, Goldstein NC, Clarke KC (2003) The spatiotemporal form of urban growth: measurement, analysis and modelling. Remote Sens Environ 86:286–302

HMGN/UNCTN (2005) Nepal millennium development goals progress report. His Majesty's Government of Nepal, Kathmandu

Hope KP, Lekorwe MH (1999) Urbanization and environment in Southern Africa: towards a managed framework for the sustainability of cities. J Environ Plan Manag 42:837–851

Hudson WD, Ramm CW (1987) Correct formulation of the kappa coefficient of agreement. Photogramm Eng Remote Sens 53:421–422

Jensen JR (2005) Introductory digital image processing: a remote sensing perspective. Prentice Hall, Upper Saddle River

Jensen JR (2007) Remote sensing of the environment: an earth resource perspective, 2nd edn. Prentice-Hall, Upper Saddle River

Kamusoko C, Aniya M (2007) Land use/cover change and landscape fragmentation analysis in the Bindura district, Zimbabwe. Land Degrad Dev 18:221–233

Kamusoko C, Masamu A, Bongo A, Munyaradzi M (2009) Rural sustainability under threat in Zimbabwe: simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Appl Geogr 29:435–447

Keuchel J, Naumann S, Heiler M, Siegmund A (2003) Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. Remote Sens Environ 86:530–541

Khoi DD, Murayama Y (2010) Forecasting areas vulnerable to forest conversion in the Tam Dao National Park Region, Vietnam. Remote Sens 2:1249–1272

Knight K (1990) Connectionist ideas and algorithms. Commun ACM 33:58–74

Leica Geosystems Geospatial Imaging (2005) Erdas field guide. Leica Geosystems Geospatial Imaging, LLC., Norcross

Li X, Yeh AGO (2004) Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landsc Urban Plann 69:335–354

Lillesand TM, Kiefer RW, Chipman JW (2008) Remote sensing and image interpretation. Wiley, New York

Lippmann RP (1987) IEEE ASSP Mag 4:4

Lwin KK, Murayama Y, Mizutani C (2012) Quantitative vs. Qualitative Geospatial Data in Spatial Modeling and Decision Making. Journal of Geographic Information Systems 4(3): (In Press)

Lwin KK, Murayama Y (2011) Mapping the human settlement of South East Asian cities using ALOS AVNIR-2, Tsukuba Geoenvironmental Sciences 7:13–17

Mesev V (2003) Urban land use uncertainty: Bayesian approaches to urban image classification. In: Mesev V (ed) Remotely sensed cities. Taylor & Francis, New York, pp 207–222

Miller RB, Small C (2003) Cities from space: potential applications of remote sensing in urban environmental research and policy. Environ Sci Pol 6:129–137

Mundia CN, Murayama Y (2011) Modeling spatial processes of urban growth in African cities: a case study of Nairobi city. Urban Geogr 31(2):259–272

Odewahn SC, Stockwell EB, Pennington RL, Humphreys RM, Zumach WA (1992) Automated star/galaxy discrimination with neural networks. Astron J 103:318–331

Prenzel B, Treitz P (2005) Comparison of function- and structure-based schemes for classification of remotely sensed data. Int J Remote Sens 26:543–561

Rakodi C (1997) The urban challenge in Africa: growth and management of its large cities. United Nations University Press, Tokyo

Richards JA, Jia X (1999) Remote sensing digital image analysis: an introduction, 3rd edn. Springer, Berlin

Rumelhart DE, McClelland JL (1988) Parallel distributed processing, 1. MIT Press, Cambridge

Seto KC, Kaufmann RK (2003) Modeling the drivers of urban land use change in the Perl River Delta, China: Integrating remote sensing with socioeconomic data. Land Econ 79:106–121

Swain PH, Davis SM (eds) (1978) Remote sensing: the quantitative approach. McGraw-Hill, New York

Thapa RB, Murayama Y (2008) Spatial structure of land use dynamics in Kathmandu Valley. In: Chen J, Jiang J, Peled A (eds) The international archives of the photogrammetry, remote sensing and spatial information sciences, Proceedings of Commission VIII, ISPRS Congress, vol XXXVII (Part-B8). ISPRS, Beijing, pp 11–16

Thapa RB, Murayama Y (2009a) Examining spatiotemporal urbanization patterns in Kathmandu valley, Nepal: Remote sensing and spatial metrics approaches. Rem Sens 1:534–556

Thapa RB, Murayama Y (2009b) Urban mapping, accuracy, & image classification: a comparison of multiple approaches in Tsukuba City, Japan. Appl Geogr 29:135–144

Thapa RB, Murayama Y, Ale S (2008) Kathmandu. Cities 25:45–57

Wang F (1990) Improving remote sensing image analysis through fuzzy information representation. Photogramm Eng Rem Sens 56:1163–1169

Weber C, Puissant A (2003) Urbanization pressure and modeling of urban growth: Example of the Tunis Metropolitan Area. Rem Sens Environ 86:341–352

Weng Y (2007) Spatiotemporal changes of landscape pattern in response to urbanization. Landsc Urban Plann 81:341–353

White RL (1996) Methods of classification. http://sundog.stsci.edu/rick/SCMA/node2.html. Accessed 15 Feb 2010

Yu X, Ng C (2006) An integrated evaluation of landscape change using remote sensing and landscape metrics: a case study of Panyu, Guangzhou. Int J Rem Sens 27:1075–1092

Yuan H, van der Wiele CF, Khorram S (2009) An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. Rem Sens 1:243–265

Zandbergen P (2008) Applications of shuttle radar topography mission elevation data. Geogr Compass 2:1404–1431

Chapter 3
Data Collection, Processing, and Applications for Geospatial Analysis

Ko Ko Lwin, Ronald C. Estoque, and Yuji Murayama

3.1 Introduction

Geospatial data collection is an important task for many spatial information users. Geospatial data collection may include field data collection, remote sensing data processing, and in-house geographical information science (GIS) data conversion. Nowadays, geospatial data are available from various sources. Among these, remote sensing data (i.e., optical, radio detection and ranging (RADAR), light detection and ranging (LIDAR), etc.) are among the primary data sources in many GIS analyzes. For example, high-resolution satellite images such as QuickBird, IKONOS, and aerial photographs are the basis for the generation of qualitative land-use maps (i.e., land-use zoning maps) and the delineation of transportation networks. Medium-resolution satellite images such as ALOS, SPOT, and Landsat TM/ETM are used in the generation of quantitative land-use maps (i.e., land cover maps) for regional-scale studies of changes in land use. The shuttle radar topography mission (SRTM) and LIDAR provide topographical characteristics for GIS analysis. Moreover, remote sensing data are important for environmental studies such as deforestation, global warming, and natural resource management. This technology captures the real-world information with various sophisticated sensors and platforms. However,

K.K. Lwin (✉) • Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kokolwin2002@yahoo.com; mura@geoenv.tsukuba.ac.jp

R.C. Estoque
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Don Mariano Marcos Memorial State University, Bacnotan, La Union, Philippines
e-mail: purplebee80@yahoo.co.uk

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_3, © Springer Japan 2012

building a GIS database is required for further geospatial analysis and mapping purposes. GIS converts the real-world information into a geodatabase in order to retrieve, analyze, and allow further geocomputations. On the other hand, field data collection is important for spatial information users in order to collect spatially distributed objects with their associated attribute information. In this chapter, we discuss geospatial data collection methods and processing, and their applications in GIS.

3.2 Geospatial Data Collection

Two approaches are presented, namely field data collection and in-house GIS data conversion. Both are frequently used in geospatial analysis and modeling.

3.2.1 Field Data Collection

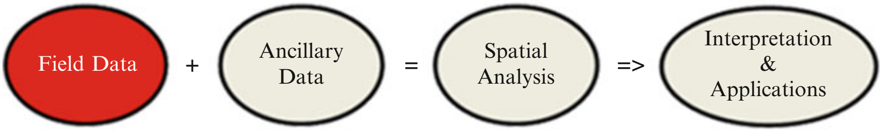
This is one spatial data collection method, and is a first-step requirement for many spatial information users such as human geographers, physical geographers, geologists, crop scientists, ecologists, etc. Human geographers may want to collect public opinions and other social activities in order to understand how social behavior changes over space and time. Geologists or physical geographers may want to collect in-situ data in order to understand overall regional geological formations and structures. Researchers or students may collect ground-truth data to validate their results.

Components of field data: Field data collection is the foundation of many spatial analysis processes. Like other spatial data, field data (Fig. 3.1) are composed of two elements, namely the coordinate information of the spatial objects and their associated attribute information. Coordinate information includes *X*, *Y*, and *Z* for the positions of spatial objects, while attribute information includes properties of those spatial objects such as the soil nitrogen contents, the names of plant or animal species, the angles of dips and strikes for each rock unit, and so on. Planning and designing attribute data are at the heart of any field data collection process, and thus it is very important that this is considered before going into the field.

Collection of coordinate information in the field: Coordinate information can be collected in several ways in the field, such as by using a global positioning system (GPS) device or GPS built-in devices, by using high-resolution satellite images, and by address matching/geocoding (conversion of addresses to *X*, *Y* coordinates) (Table 3.1).

Field data collection methods: Along with recent advances in modern wireless communication and Internet technologies, and mobile computational devices, nowadays field data collection can be conducted in a handy and timely manner. Many methods have been developed for field data collection, ranging from personal field data collection to automatic real-time field data collection using GPS, personal digital

Field Data and Spatial Analysis



Field Data	Ancillary Data	Spatial Analysis	Interpretation & Applications
Plants	Census	Spatial analysis	Finding population mean center
Soils	Roads	Spatial modeling	Cluster or disperse
Rocks	Land use	Spatial statistics	Spatial relationship
Weather	DEM	Spatial indices	...
Land use/cover			
Ground truths			

Field Data Structure

Location
X,Y,Z

Attribute
Geologists: rock type, dip, strike, bearing, lithology, ...
Ecologists: plant, insect species, count, ...
Climatologists: temperature, wing speed, direction, land use, ...
Human Geographers: sex, race, religious, income, ...

X	Y	TYPE	VALUE_01	REMARK	ATTACHMENT	IMAGES
418336.68	3997566.4	Temperature	32.6,12.2,194,bar	Dried crop land	1	AW020726.jpg
418556.28	3996879.1	Temperature	35.2,16.2,180,bar	Agricultural test center	1	48f423e1.jpg
419217.74	3996271.3	Temperature	28.3,13.2,270,wat	Near the pond	1	pine-tre.jpg
418943.9	3994777.0	Temperature	28.7,5.2,202,park	Kasuga koren	1	park.jpg
419490.26	3996394.9	Temperature	26.3,3.2,280,for	Forest	1	TN_fores.jpg
418308.23	3994644.1	Temperature	30.2,5.6,230,grs	This is grass land	1	179465_3.jpg
418837.14	3994496.6	Temperature	31.5,3.2,183,urb	Urban area	1	181_8141.jpg
418874.66	3996773.6	Temperature	34.2,10.3,250,car	Car parking	1	parking_ .jpg
419330.35	3995495.2	Temperature	26.3,5.4,180,wat	Pond	1	pond1.jpg
419566.91	3993898.3	Temperature	30.2,7.4,190,urb	Built-up area	1	photo_ka.jpg

Fig. 3.1 Importance of field data collection in geospatial analysis and elements of field data

assistants (PDAs), and other mobile computational devices such as ultra-mobile personal computers (UMPC), smart phones, and Netbooks. In the following subsections, two field data collection methods are discussed, namely personal field data collection using an ultra-mobile PC (UMPC) or a Netbook computer, and real-time field data collection using a mobile phone. The latter is used to collect field data through a Post Office Protocol 3 (POP3) mail server and a centralized geodatabase, either individually or group-based. It is ideal for individual field data collection.

3.2.1.1 Personal Field Data Collection

Recent innovations in computing, networking, and Internet technologies enable GIS field users to collect, store, and analyze in a handy and mobile manner. PDAs are hand-held computers that were originally designed as personal data organizers, but became more versatile over the years. A PDA can be used as a clock, as a calendar, for accessing the Internet, for sending and receiving e-mails, for working on spreadsheets,

Table 3.1 Coordinates collection methods and their advantages and disadvantages for various applications

Method	Advantages	Disadvantages
Using a GPS or GPS built-in device	• Good for rural areas where open spaces are abundant and landmarks (road intersections, building shapes, etc.) are absent • Small and handy • Can be integrated with other computational devices (e.g., Notebook, Netbook, etc.) • Good for geologists, archaeologists, ecologists, etc.	• Landscape-dependent (accuracy varies with type of landscape) (open space, 4 m; semi-open space, 10 m; closed space, 17 m in the case of a hand-held Garmin GPS)
Using a base map of high-resolution satellite images	• Good for urban area studies where the landmarks are clear (road intersections, building shapes, etc.) • Landscape-independent	• Requires prior knowledge of that area • Requires software to use it • Costly[a] • Requires additional time for data processing[a] • Needs more storage space (large file size)[a]
Geocoding and address matching	• Good for a large amount of data conversion such as patient addresses, customer addresses, … • Timely manner • Good for finding spatial relationships between humans and the environment, health, disease dispersion, etc.	• Requires a national-scale geodatabase such as the National Spatial Data Infrastructure (NSDI) • Accuracy depends on the available dataset, such as the level of the street or the level of a parcel unit • Not suitable for developing countries where fine-scale GIS data are absent

[a]If the GIS application can be integrated with Google Maps or Microsoft Bing maps through the Web map service (WMS) in an application, these disadvantages can be ignored. For example, Google WMS is a WMS interface to Google Maps. This allows users to access Google map data from GIS applications via an HTTP interface. It can provide Google map image data to any GIS application that can use a WMS service for raster data

and for using a word processor. However, PDAs lack the fully blown infrastructure of a wireless broadband network and have a limited screen resolution (typically 240×320). UMPCs (typically with a screen resolution of 1024×600 wide screen) began as a joint development exercise by Microsoft, Intel, Samsung, and others. UMPCs are able to run any software that has been written for the Windows XP platform. UMPCs can also feature GPS devices, Wi-Fi, and Ethernet. There has been a revolution in GPS over the last few years as the cost of receivers has decreased and accuracy has improved. GPS has become a critical tool for spatial information users in a wide range of application fields. Owing to the characteristics listed above, the mobile GIS is rapidly gaining popularity and effectiveness among spatial information users. A mobile GIS is also interdisciplinary. Nowadays, the focus of much leading-edge research in geography is interdisciplinary (integrating two or more academic disciplines), and hence is not limited to scientific investigation only, but can also be extended to applied real-world problem solving (so-called normative uses) in a time- and cost-effective way. Google Maps provides high-resolution satellite images for almost all urban areas around the world. The spatial resolution is good enough to collect reference ground control points (GCPs) in urban areas owing to the nature of distinguishable landscapes and their associated features, such as road intersections, building shapes, etc. Urban area field surveys such as household surveying, road condition inspection, hydrant inspection and mapping, damage investigation (in the case of a disaster), public health surveys, and the collection of public facilities are important to local and city planners to help in effective urban planning.

Initially, UM-FieldGIS (Lwin and Murayama 2007) was intended for use in student field survey projects which did not require any server-side installation. UM-FieldGIS is a Windows-based GIS program which allows the user to collect, store, and integrate information into a current GIS system in a timely manner. UM-FieldGIS field survey files are based on the Microsoft access database format (.mdb file extension), a technology that is also used in ESRI personal geodatabase (.mdb) (Fig. 3.2). However, ESRI personal geodatabases are more complex than MS access databases as they handle both the database management system (DBMS) and the geographical features, as well as their associated topographical relationships. Under the UM-FieldGIS, the field user can collect geographical positions from the embedded Google Maps API (online network connection), or a pre-installed map (PIM) (map-based mobile GIS), or an attached universal serial bus (USB) GPS, or a built-in GPS. Field users are able to create their own survey items (field names) and choose a multimedia attachment capability. UM-FieldGIS supports general GIS functions such as map zooming, scrolling, querying attributes, arranging map layers, searching by attribute name, and changing the properties of map views under the PIM mode. Finally, the surveyed data can be exported into the ESRI shape file format, or directly imported by the ArcGIS software.

The UM-FieldGIS graphic user interface (Fig. 3.3) is especially designed for UMPC wide-screen resolution (1024×600) monitors and is featured on the tabbed dialog box class. Functionally, the UM-FieldGIS can be divided into two, the database module and the GIS module. All functions are grouped by tabs. The user can easily switch between functions by clicking the appropriate tab without necessarily opening multiple windows. This is especially advantageous for a desktop

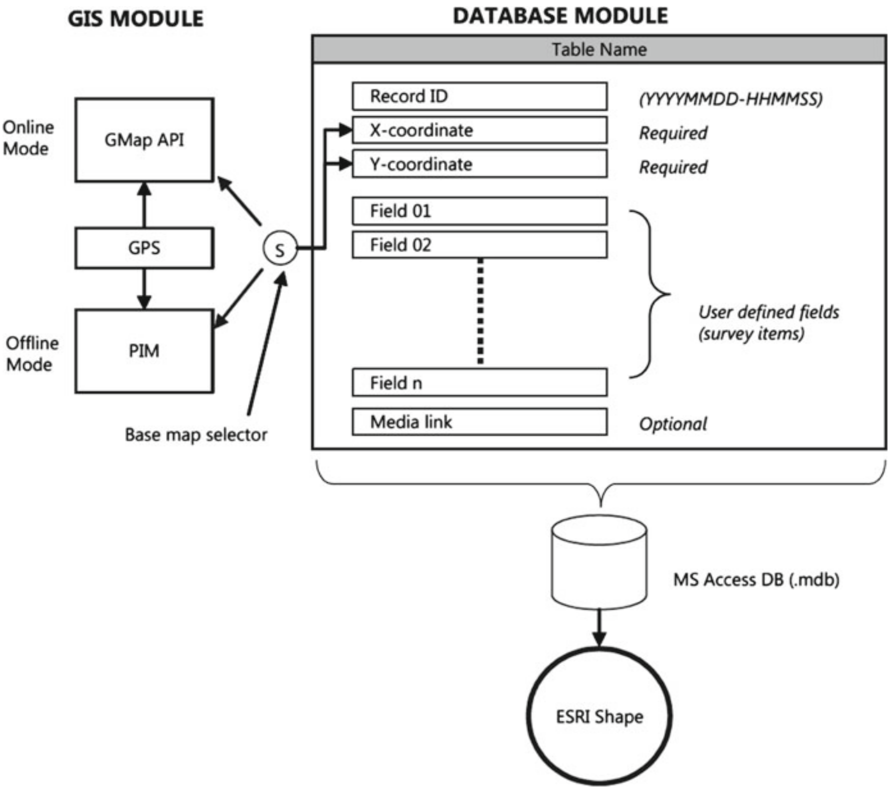


Fig. 3.2 Program Design of *UM-FieldGIS.PIM*, Pre-installed Map (supported formats: GeoTIFF and ESRI Shape); *GMap API*, Google Maps Application Program Interface; Download URL: http://giswin.geo.tsukuba.ac.jp/sis/en/gis_software.html

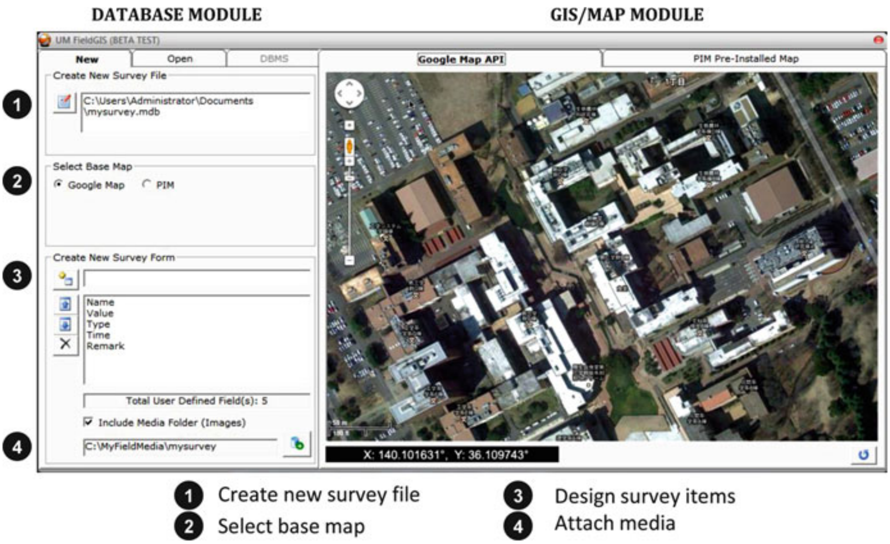


Fig. 3.3 Graphical user interface of the UM-FieldGIS

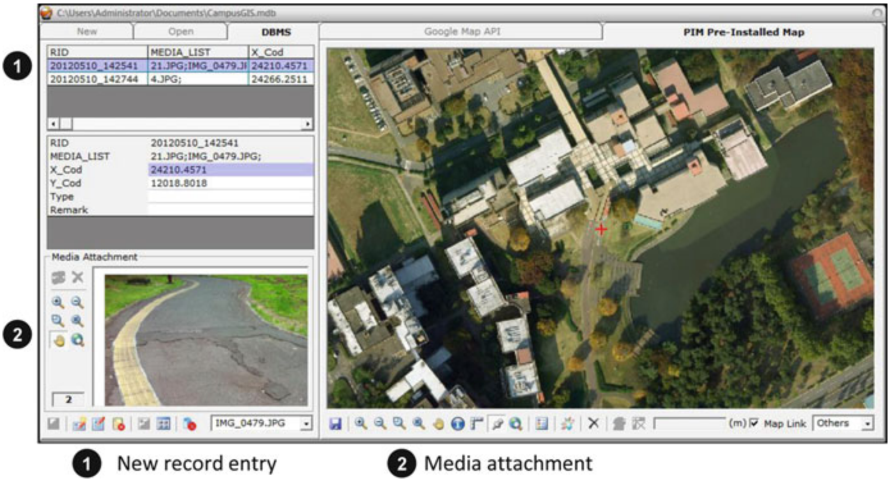


Fig. 3.4 Add new record

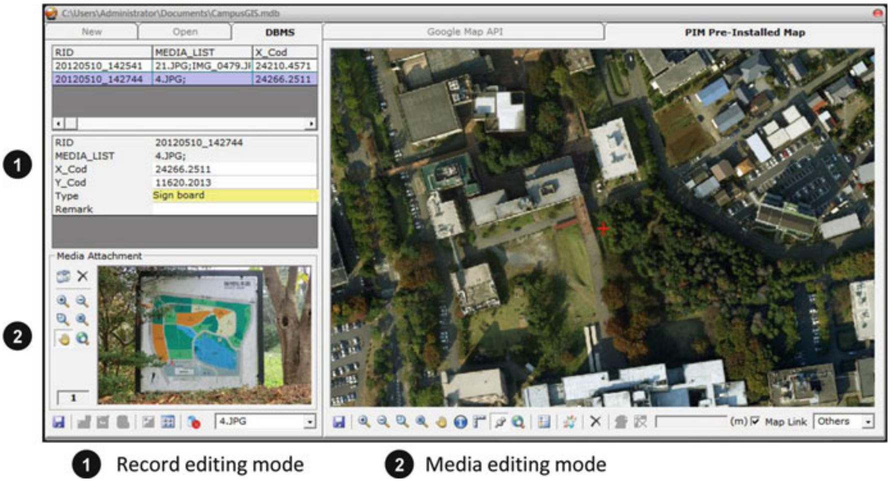


Fig. 3.5 Edit record

with a limited screen display. The database module enables the user to create a new field survey file in Microsoft access database format (.mdb), or open an existing survey file to analyze or to append new information during the surveying process.

The UM-FieldGIS allows users to create their own survey form (survey items) and add a media folder for image attachments in each record (Fig. 3.4). Records can be added, deleted, and edited using the add new/edit mode tab. Collected records can be viewed in a tabular, record, or media view (Fig. 3.5).

The UM-FieldGIS is suitable for field data collection in urban areas where wireless access services such as Wi-Fi is available. However, it can still be used in PIM mode by using maps or high-resolution satellite images without Internet access.

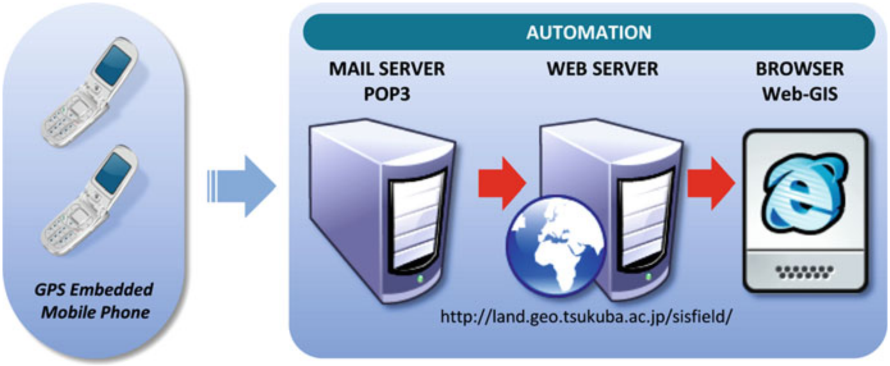


Fig. 3.6 A centralized geodatabase and mobile field data collection

3.2.1.2 Real-Time Field Data Collection (Centralized Geodatabase)

UMPC or Netbook computers and wireless Internet access services are sometimes costly and are not suitable for all users. We need to find other methods of collecting field data in a handy and timely manner at low cost, such as by using a personal mobile phone. Figure 3.6 shows an overview of the real-time field data collection system. Basically, this system consists of two sub-modules, namely the client module and the automation module. The client module contains only a GPS-embedded mobile phone or a GPS-plus mobile phone. All the functions of the data injection and format conversion processes are performed automatically within the automation module. Finally, the real-time results can be viewed through a Web browser by providing Web-GIS.

How It Works

Real-time field data collection (a centralized geodatabase) utilizes GPS-embedded mobile phones, which typically support additional services such as the short messaging service (SMS), the multimessaging service (MMS), e-mail and Internet access, and short-range wireless (infrared or Bluetooth) communications, as well as business and gaming applications, and photography. Users are required to type a predefined text format for collecting the data. For example, the user needs to add a "/" character between fields, and add a "," between attribute values (Fig. 3.7). This text message is then sent using a predefined mail address and subject. The user can also attach as many photographs as needed. This text message is read by the POP3 mail server, converted into a GIS dataset, and then injected into a centralized geodatabase at specific time intervals within the automation process. The centralized geodatabase is composed of aerial images, other ancillary datasets, and the injected data (survey data). End users can download and visualize the survey data in ESRI-shape file format through a Web browser for further analysis.

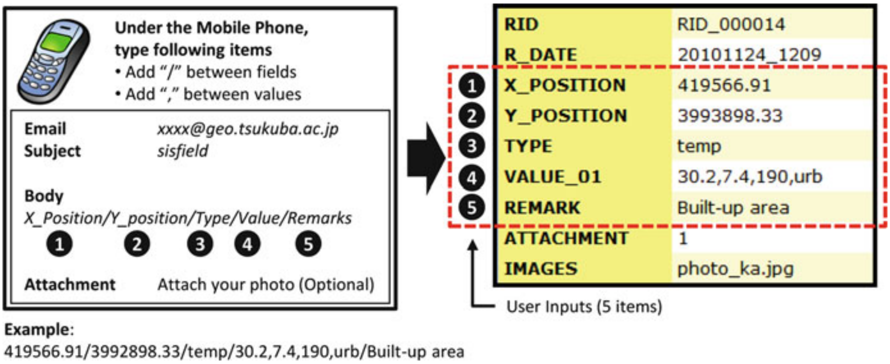


Fig. 3.7 Format conversion between a text message and an attribute table in the geodatabase (Lwin and Murayama 2011b)

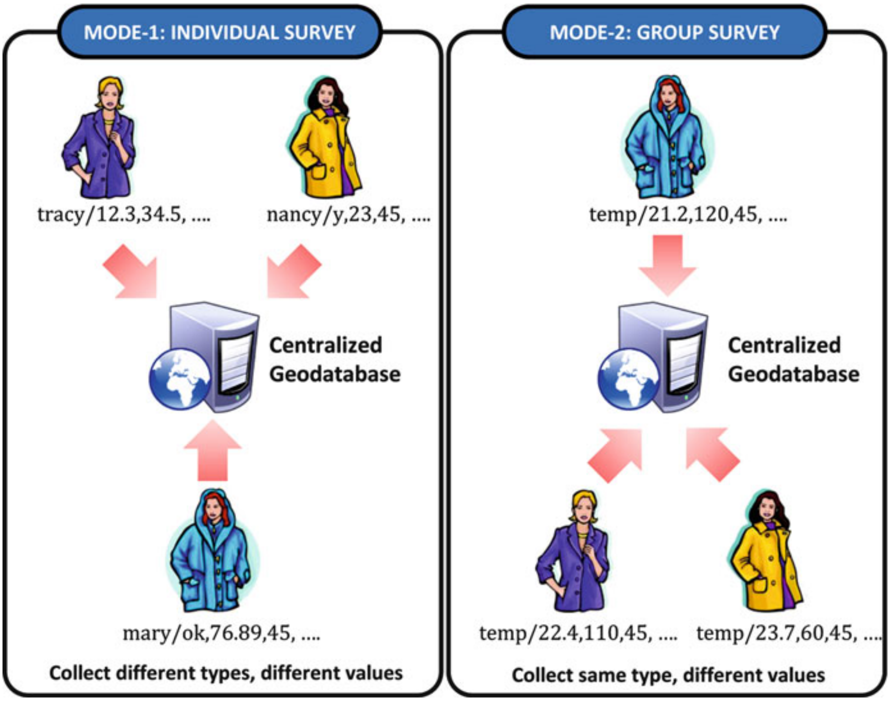


Fig. 3.8 Modes of survey: individual and group (Lwin and Murayama 2011b)

This filed data collection method can be implemented through an individual or a group survey by changing the "type" field (Fig. 3.8). The individual survey mode is ideal for users who collect field data for their own specific purpose, while the group survey mode is ideal for real-time collection of information such as surface temperature, wind speed, wind direction, damage information, etc. For the individual survey mode, the “type” field may be the user’s initial name. Later, the users would be able to extract their own data by using this field. For the group survey mode, the “type” field may be a category which is being surveyed, such as temperature, land-use type, rock or soil properties, etc.

The overall system is built on Microsoft ASP.NET with an AJAX extension and VDS technologies (Web mapping components for ASP.NET). ASP.NET is a Web application framework marketed by Microsoft that programmers can use to build dynamic Websites, Web applications, and XML Web services. AJAX (shorthand for asynchronous JavaScript and XML) is a group of interrelated Web-development techniques used on the client side to create interactive Web applications. With AJAX, Web applications can retrieve data from the server asynchronously in the background without interfering with the display and behavior of the existing page. The use of AJAX techniques has led to an increase in interactive and dynamic interfaces on Web pages. The AspMap for .NET from VDS technologies is a set of high-performance Web-mapping components and controls for embedding maps in ASP.NET applications (Web forms).

This field data collection method has been introduced to the students of the University of Tsukuba, Japan, during their field survey course, which is part of the university campus GIS project. Under the campus GIS project, individual students are required to collect or report illegal bike or motorbike parking places, illegal waste disposal site locations, and man-made footprints which are caused by people who walk on the grasslands or who are passing between trees instead of using legal paths. Later, this information is used by the university administrators to maintain the campus landscape and manage student facilities. A group survey was also conducted to collect environmental data such as surface temperature, wind speed, wind direction, etc., on a real-time basis. In this case study, 4 faculty members and 16 students from the University of Tsukuba, Japan, and 2 faculty members and 9 students from the South China Normal University, China, participated.

Planning ahead is important for adequate and successful field data collection. Spatial planning and sampling design include setting where and what attribute information is to be collected. Sometimes, it is difficult or impossible to collect again after the field work has been done once. In this project, we assigned the survey area to student groups based on administrative units. We also demonstrated the handling of GPS and other field survey instruments. Students were required to send field survey data by using their GPS-embedded mobile phone or by reading the coordinates from a Garmin hand-held GPS. During the field work, we monitored their status on a Netbook computer with wireless Internet access (Figs. 3.9 and 3.10). We also advised the students through mobile phone communication.

After the field work, students were required to download the survey data through Web-GIS and open it in ESRI ArcMap in the laboratory. This process includes downloading the data, importing it in ArcGIS, formatting the data, and visualizing it in ArcMap. We used the following Visual Basic (VB) scripts to format the comma-separated values into attribute fields (Fig. 3.11). String substitution was also carried out by VB Script to replace the short text with full text, such as “urb” to “Urban.” This is because some students collected and recorded the data using short text

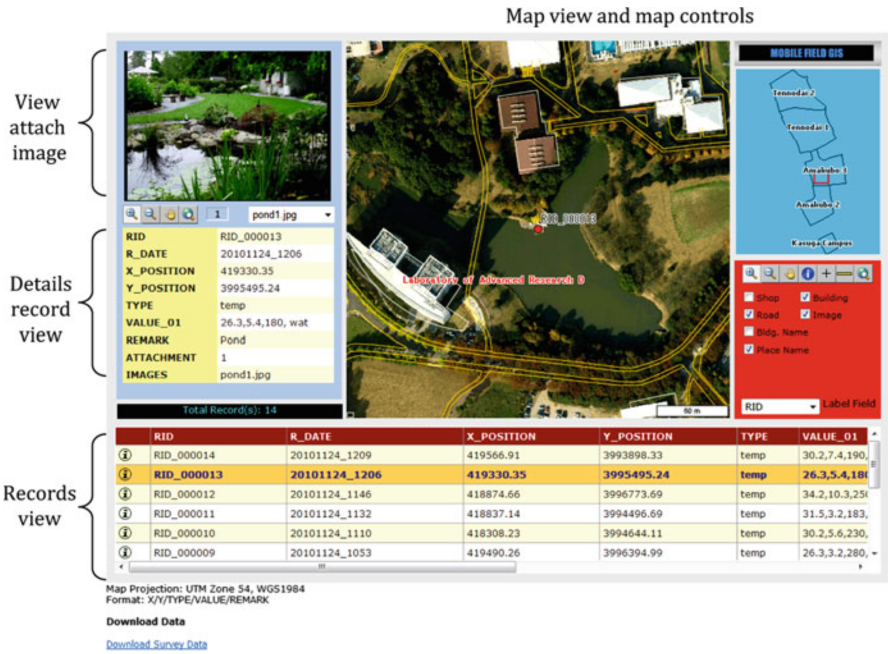


Fig. 3.9 Acquired real-time information with media attachment (Lwin and Murayama 2011b)

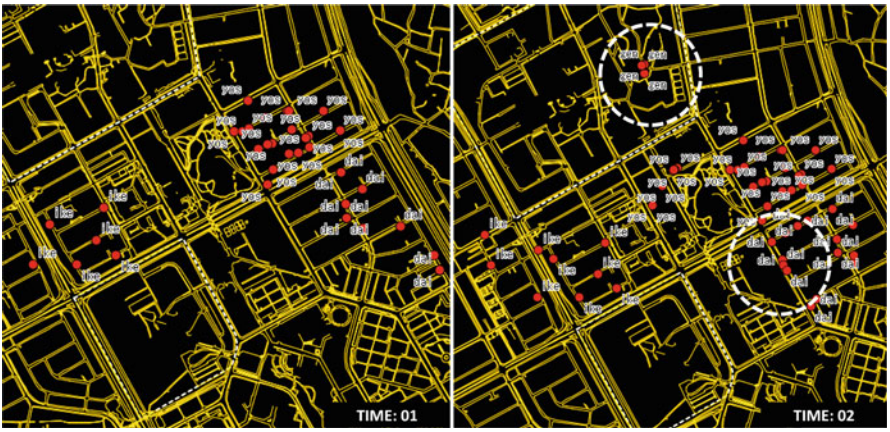


Fig. 3.10 Real-time data injection (Lwin and Murayama 2011b)

messages to reduce the typing time and errors. Furthermore, students were also able to specify their names in the "Type" field.

Every year, a field survey is conducted to collect information about public facilities such as bicycle stands and capacity, garbage boxes and types, car parking lots, sidewalk conditions, illegal garbage places, and other environmental data. Figure 3.12 shows the visualization of student survey data in the ArcGIS software.

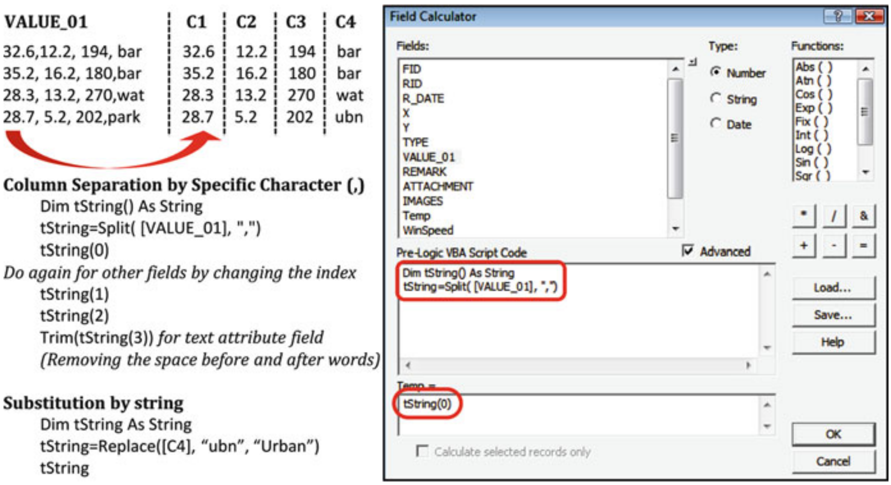


Fig. 3.11 Formatting the data in ArcGIS

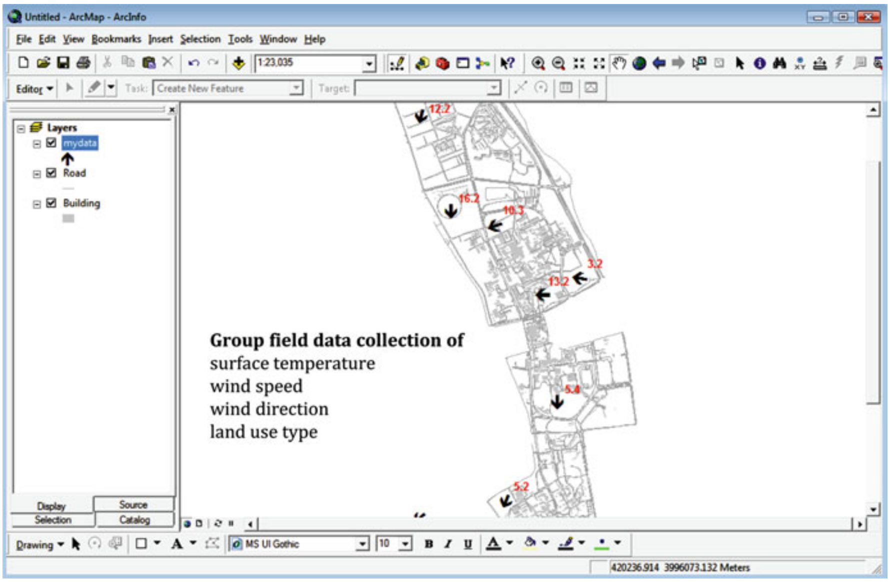


Fig. 3.12 Mapping the surface temperature, wind speed, and wind direction in ArcMap based on a student group survey data

3.2.2 In-House GIS Data Conversion

The two main GIS data sources include digital and non-digital sources. Digital sources include information captured through remote sensing and surveying (field data), while non-digital sources include, but are not limited to, paper maps, which are usually digitalized through scanning and digitization (Sheldon 2007). This sub-section discusses the "in-house data collection" method. Generally, this method involves the digitalization of available paper maps. Map sheets comprise one of the most widely available and familiar sources of spatial data (Malczewski 2004). The immediate outputs of this method are digitized GIS vector data, which include points, lines, and polygons. However, the production of these usually involves the processes of georeferencing and digitizing, which are also preceded by scanning when the digitizing process is to be done on-screen.

Scanning: Prior to on-screen digitizing, paper maps have to be integrated into the GIS database by converting them into digital format. The process of such conversion is known as scanning. Through scanning, map features, including texts and symbols, are automatically captured as individual cells or pixels and an automated image is produced. These features in raster format are then vectorized through tracing or on-screen digitizing. Generally, in order to have a good source image in the digitizing process, a scanner needs to have a good resolution and, depending on the specific purpose, has to be large enough to accommodate the map sheets being scanned.

Georeferencing: Basically, the process of projecting image data onto a plane and making it conform to a map projection system is called rectification. However, since scanning produces images that are already planar, rectification is no longer required unless there is distortion in the image (Leica Geosystems 2005). In this instance, such scanned images only need to be georeferenced. Georeferencing refers to the process of assigning map coordinates to image data. As explained in the ERDAS imagine field guide (Leica Geosystems 2005), "the image data may already be projected onto the desired plane, but not yet referenced to the proper coordinate system. Rectification, by definition, involves georeferencing, since all map projection systems are associated with map coordinates. Image-to-image registration involves georeferencing only if the reference image is already georeferenced. Georeferencing by itself involves changing only the map coordinate information in the image file. The grid of the image does not change."

Digitizing: Before the on-screen digitizing technique became available, table digitizing was the acceptable method for creating GIS vector data. However, aside from being considered as a time-consuming process, several challenges have also been experienced with this method, such as difficulty in using the digitizing puck, tablet malfunctions, source materials changing size, registration problems, edge-matching complexity, and more (Sheldon 2007).

Today, with the availability of more advanced digitizing methods like on-screen digitizing, it is possible to derive GIS vector data from a sheet map using a desktop computer (Fig. 3.13). On-screen digitizing only requires that the sheet map is scanned and properly georeferenced. With this method, which is sometimes referred to as "heads-up" digitizing, the scanned and georeferenced source map or photograph is displayed on-screen, and features are digitized using a standard mouse (Eastman 2006). This advancement in technology allows users to "zoom in" on images as much as is needed, digitize using a computer mouse, and edge-match more easily for faster map creation (Sheldon 2007).

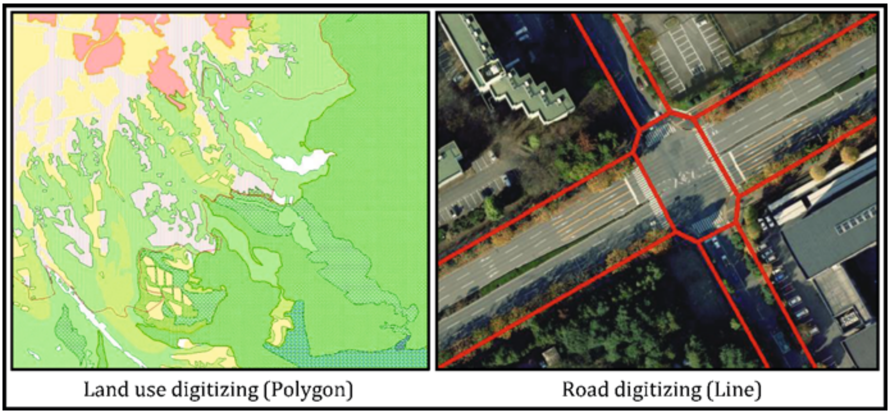


Fig. 3.13 On-screen digitizing for GIS data conversion

3.3 Geospatial Data Processing and Applications

Geospatial data processing is the heart of the task in many GIS analyzes. It is necessary to understand digital image processing, database design and construction, and GIS analytical functions. Geospatial data can be grouped into raster and vector formats. Raster data are mainly derived from remote-sensing data, and vector data are mainly constructed in GIS.

3.3.1 Raster Data Processing and Applications in Geospatial Analysis

Many remote-sensing data used in GIS are in raster format, such as land-use land cover (quantitative), the normalized difference vegetation index (NDVI), the satellite-derived digital elevation model (DEM), and surface temperature, since the remote-sensing technology captures the real-world information pixel by pixel, which is known as spatial resolution with sophisticated sensors. This section discusses some of the raster data commonly used in geospatial analysis, such as the quantitative land-use land cover and the NDVI.

3.3.1.1 Land-Use Land Cover

Quantitative land use/cover maps are commonly derived from medium-resolution satellite images using a method known as multispectral classification. In developing countries, land use/cover maps derived from satellite images are a cost- and labor-effective measure for updating an existing land use/cover map database and for

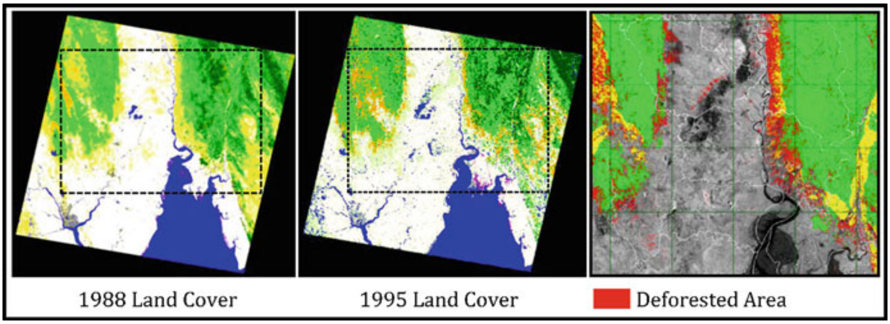


Fig. 3.14 Delineation of deforested areas using Landsat TM images

managing the country's natural resources (Lwin and Shibasaki 1997; Nunes and Auge 1999). The applications of land use/cover maps in GIS are manifold, e.g., land-use change modeling, monitoring of deforestation, natural resource management, and hydrological modeling. Figure 3.14, for example, shows the combination of two different spatial and temporal resolution satellite images, Landsat ETM and NOAA AVHRR data, to monitor the annual deforestation rates and deforested areas in Myanmar (Lwin and Shibasaki 1998).

3.3.1.2 Normalized Difference Vegetation Index (NDVI)

The normalized difference vegetation index (NDVI) can be derived from normalizing two spectral bands, the infrared and red bands (NDVI = RED – IR/RED + IR). Chapter 16 presents the use of advanced land observing satellite (ALOS) image-derived NDVI data to model urban green walkability space. Furthermore, NDVI data can be used to monitor annual deforestation rates by using 10-day composite NOAA AVHRR 1-km data for regional scale studies (Fig. 3.15).

Figure 3.16 shows the surface temperature derived from Landsat TM band 6 (thermal band). For example, Lwin and Murayama (2010b) detected the urban thermal fringes from Landsat ETM-derived surface temperatures using "focal statistic analysis" in GIS to identify surface temperature variations (i.e., the heat island effect) inside Tsukuba city, in order to take further action for eco-city planning.

3.3.2 Vector Data Processing and Applications in Geospatial Analysis

Vector data are composed of points, lines, and polygons. Vector data can be directly purchased from map vendors, or generated by in-house GIS data conversion methods such as scanning, georeferencing, and digitizing.

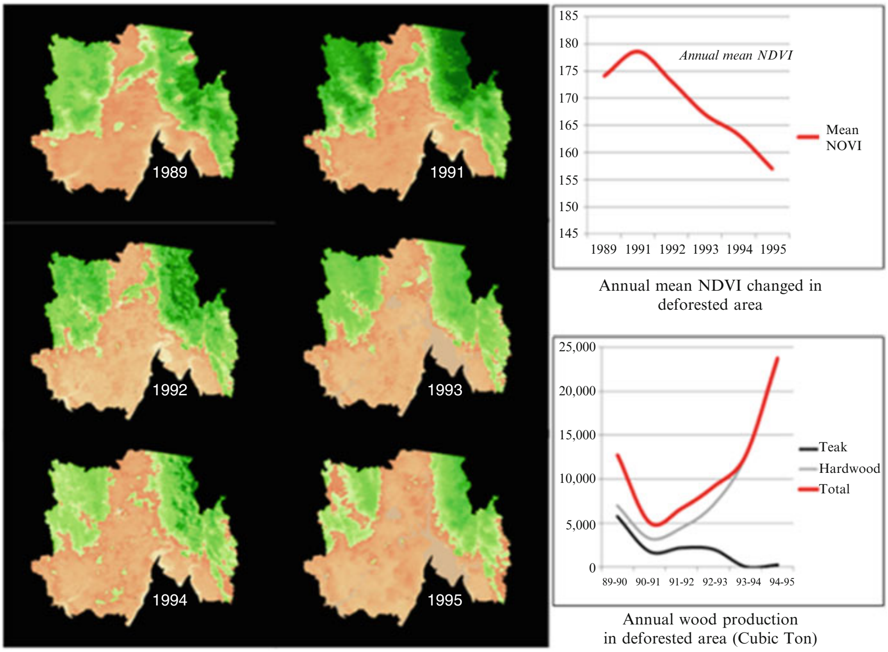


Fig. 3.15 Monitoring the annual deforestation process using NOAA AVHRR 10-day composite NDVI data (a case study in Myanmar)

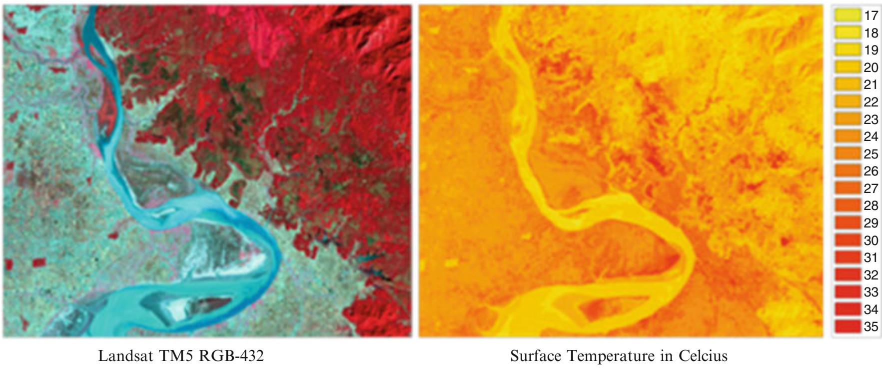


Fig. 3.16 Surface temperatures derived from Landsat TM5

3.3.2.1 LiDAR Point Data Processing and Applications in Geospatial Analysis

Point data are fundamental elements in vector data. Point data represent bus stops, facility locations, field survey data of soil and rock properties, surface temperature, etc. This sub-section considers the use of LIDAR point cloud data in generating

DEMs, digital surface models (DSM), digital terrain models (DTM), and digital height models (DHM). The terms DSM, DTM, and DHM are generally suitable for high-resolution digital elevation data, since such data can distinguish the heights between buildings, trees, and other built-up objects, while DEM is generally suitable for coarse spatial resolution elevation data (i.e., 30 m, 90 m, 1 km).

Traditionally, stereo-image matching is a standard photogrammetric technique to generate a DSM. However, this technique is good only for a smooth open terrain surface. The quality of a DSM in built-up areas is poor owing to occlusions and height discontinuities (Haala and Brenner 1999). LiDAR techniques have been studied and utilized since the early 1960s, but appear to have become more prominent in the past few years. LiDAR has found applications in a wide variety of fields of study, including atmospheric science, bathymetric data collection, law enforcement, telecommunications, and even steel production (Maune et al. 2000). Because LiDAR operates at much shorter wavelengths, it has a higher accuracy and resolution than microwave radar (Jelalian 1992).

Choosing the appropriate surface-generating method for DSM and DTM is important in LiDAR data processing, since surface height information is collected as points. Figure 3.17 shows the detailed procedure for generating a DHM and digital volume model (DVM). In this process, both DSM and DTM point features are converted into a triangulated irregular network (TIN) model (i.e., TIN–DSM and TIN–DTM). Using ArcGIS software, the TIN process allows users to convert multiple scenes at one time. This reduces the time for mosaicing. Moreover, the TIN process is faster than other interpolation processes such as IDW, SPLINE, and kriging. Each TIN–DSM and TIN–DTM is converted into a raster format, setting the spatial resolution to 0.5 m. The DTM is subtracted from the DSM raster layers to achieve the DHM. For this DHM raster layer to be converted into a DVM, it is multiplied by the cell surface area (i.e., 0.25 m^2).

The possible uses of elevation data in GIS are numerous, e.g., the identification of river flow directions and the delineation of catchment areas (river basin or watershed) in hydrology, soil erosion modeling, building height extraction for telecommunications, and other 3D applications. DVM is also used for building population estimates (Lwin and Murayama 2010a) for microscale population data analysis and 3D visualizations of urban landscapes (Fig. 3.18).

3.4 Conclusion

Perhaps the most exciting area of computer system development continues to be in hand-held devices such as PDA, UMPC, Netbooks, and smart phones. A smart phone is a mobile phone that offers more advanced computing ability and connectivity than a contemporary feature phone. They are noticeably more efficient in form factor (size, shape, weight, etc.), chip type, internal storage capacity, battery life, and operating system compared with desktop computers. Along with hardware developments, the operating systems used in smart phones are becoming more and

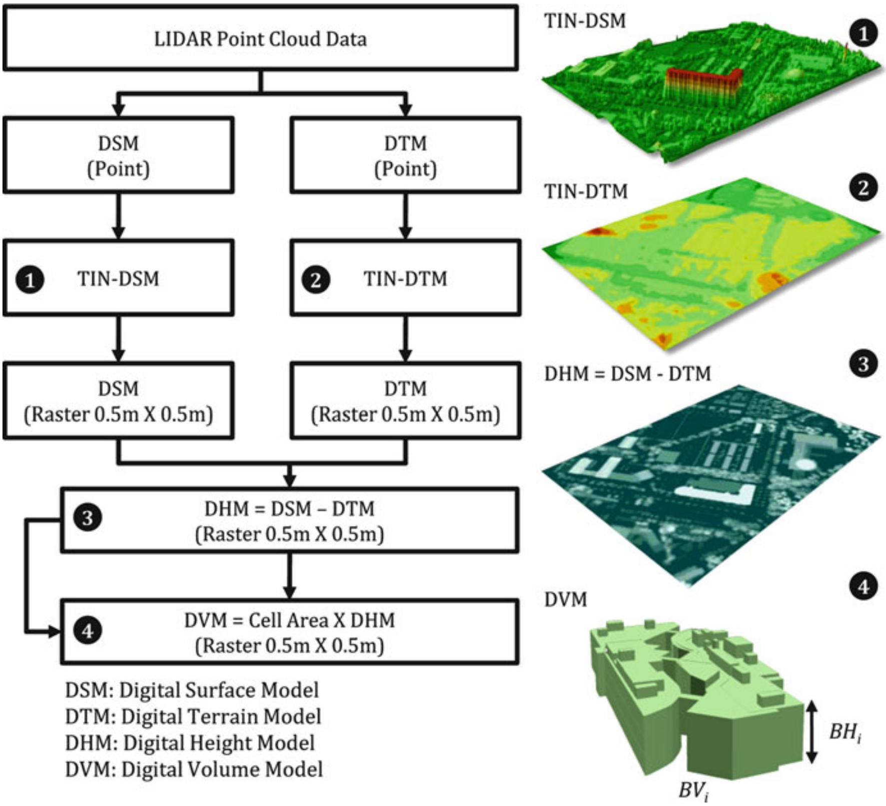


Fig. 3.17 LiDAR data processing work flow (modified from Lwin and Murayama 2010a)

more compact and functional, e.g., iPhone (Apple Inc.) and Android (Google). Computer scientists at the University of Washington have used Android, the open-source mobile operating system championed by Google, to turn a cell phone into a versatile data-collection device. They collected data on deforested areas, and instantly submitted that information to the global environmental database (University of Washington 2009). In the meantime, the number of cellular mobile phone subscribers world-wide is increasing year by year. According to the International Telecommunication Union's (ITU) 2010 report, by the end of 2009, there were an estimated 4.6 billion mobile cellular subscriptions, corresponding to 67 for every 100 inhabitants globally. Recently, a couple of studies showed field data collection with mobile phones in both the educational and industrial sectors (Mourão and Okada 2010; Moe et al. 2004).

On the other hand, the increasing popularity of the Internet and user-friendly Web-based GIS applications such as Google Maps/Earth and Microsoft Bing maps have made GIS an integral part of life today for finding the nearest facilities, driving routes, and so on. For example, in Tsukuba City, Japan, local residents and "green"

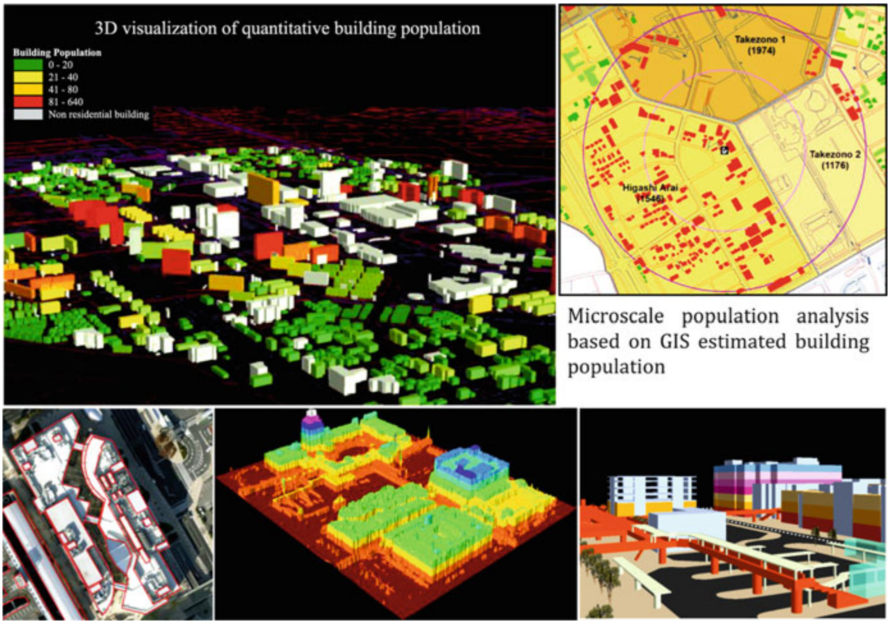


Fig. 3.18 Building population estimates from DVM (*upper*), and 3D visualizations of urban landscapes by the combination of building footprints and LIDAR height data (*below*)

exercise takers can find the shortest or greenest route between stops by using their smart phone while walking along the street and accessing the eco-friendly walk score calculator Web-based GIS (Lwin and Murayama 2011a). However, PDAs, Netbooks and smart phones are sometimes considered to be cost-intensive, including both device and wireless access service charges, and hence not suitable for use in student field survey projects. Moreover, mobile field computing environments vary widely, but generally offer extremely limited computing resources, visual display, and bandwidth relative to the usual resources required for distributed geospatial data (Nusser et al. 2003). Nevertheless, geospatial data collection, processing, and analysis tasks are important in GIS. Proper data collection and conversion are required to support the geospatial analysis that is vital for decision-making.

References

Eastman JR (2006) IDRISI Andes tutorial. Clark University, Worcester

Geosystems L (2005) ERDAS field guide. Leica Geosystems Geospatial Imaging, LLC, Norcross

Haala N, Brenner C (1999) Extraction of buildings and trees in urban environments. ISPRS J Photogramm Remote Sens 54:130–137

Jelalian A (1992) Laser radar systems. Artech House, Boston

Lwin KK, Murayama Y (2007) Personal field data collection by UM-FieldGIS (Ultra Mobile PC and Embedded Google Map API). In: 16th Papers and proceedings of the GIS Association of Japan (GISA), Hokkaido University, Sapporo, pp 165–170
Lwin KK, Murayama Y (2010a) Development of GIS tool for dasymetric mapping. Int J Geoinformatics 6:11–18
Lwin KK, Murayama Y (2010b) Urban thermal fringes detection and environmental quality assessment: a case study of Tsukuba City. Tsukuba Geoenviron Sci 6:23–27
Lwin KK, Murayama Y (2011a) Modelling of urban green space walkability: eco-friendly walk score calculator. Comput Environ Urban Syst 35:408–420
Lwin KK, Murayama Y (2011b) Web-based GIS system for real-time field data collection using personal mobile phone. J Geogr Inform Syst 3:382–389
Lwin KK, Shibasaki R (1997) Land use land cover classifications and monitoring of Myanmar using remote sensing data and GIS. In: 18th Asian Conference on Remote Sensing (ACRS), Kuala Lumpur, Malaysia
Lwin KK, Shibasaki R (1998) Monitoring and analysis of deforestation process using remote sensing and GIS: a case study in Myanmar. In: 19th Asian Conference on Remote Sensing (ACRS), Manila, Philippines
Malczewski J (2004) GIS-based land-use suitability analysis: a critical overview. Prog Plan 62:3–65
Maune D, Daniel C, Damron J (2000) LIDAR and IFSAR: Pitfalls and opportunities for our future. Proceedings of the American Society for Photogrammetry and Remote Sensing Conference, Washington, DC
Moe K, Dwolatzky B, Olst R (2004) Designing a usable mobile application for field data collection. IEEE AFRICON pp 1187–1192
Mourão S, Okada K (2010) Mobile phone as a tool for data collection in field research. World Acad Sci Eng Tech 70:222–226
Nunes C, Auge JI (eds) (1999) Land-use and land-cover implementation strategy. IGBP, Stockholm
Nusser SM, Miller LL, Clarke K, Goodchild MF (2003) Geospatial IT for mobile field data collection. Commun ACM (Association for Computing Machinery) 46:64–65
Sheldon M (2007) GIS technology trends in scanning and plotting: what's new and what's next. Direction Magazine, Illinois
University of Washington (2009) Cell phones become handheld tools for global development. Science Daily. http://www.sciencedaily.com/releases/2009/10/091029141249.htm. Accessed 9 Nov 2009

Part II
Geospatial Theories and Methods

Chapter 4
Fuzzy Set Theory in Geospatial Analysis

Yaolong Zhao

4.1 Introduction

Geographical information systems (GIS) are designed to store, retrieve, manipulate, analyze, and map geographical data. Since the 1960s when R.F. Tomlinson first presented the GIS, this field has mainly focused on the construction of the systems, the improvement of system functions, and the extension of its application to other disciplines. The research contents have played an important role in providing spatial decision-making support for both governments and the public, and have also promoted the formation and development of the discipline of Geographic Information Science (Goodchild 1992). However, with the extension and deepening of applications, users began to doubt the results of spatial analysis using GIS (Doucette and Paresi 2000; Morrison 1995; Östman 1997; Stefanakis et al. 1999). The raw material for GIS (i.e., the original data imported into GIS) inevitably always contains errors (Shi et al. 2002). Data models used in GIS to describe the real world are just approximations to objective reality. In addition, all kinds of spatial operations and processing approaches may bring new errors and uncertainties into the production of spatial analysis. Most existing designs of GIS software are based on the hypothesis that no errors exist in geographic entities and their spatial relationships. Generally, GIS can only deal with determinate spatial entities and their relationships. However, using a GIS designed to deal with determinate data for uncertain data will bring about problems, and the results cannot satisfy the users' needs (Shi et al. 2002). As the outputs of GIS play an important role in spatial decision-making support, users began to be

Y. Zhao (✉)
School of Geography, South China Normal University,
Guangzhou, People's Republic of China

Formerly in Division of Spatial Information Science, University of Tsukuba,
Tsukuba, Ibaraki, Japan
e-mail: yaolong@scnu.edu.cn

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_4,

concerned about the quality of spatial data in GIS. This undoubtedly made many scholars think about the field of GIS spatial data quality control (Mowrer and Congalton 2000; Östman 1997; Shi et al. 2002).

The acquisition of spatial data in GIS primarily relies on surveying and geographical investigation. Matured surveying errors and data processing methods (namely surveying adjustments) which are based on probability and statistical theories have been introduced into the field of spatial data quality control. Thus, a set of theoretical systems on spatial data quality control methods gradually came into existence (Goodchild and Dubuc 1987; Goodchild and Gopal 1989; Heuvelink et al. 1989; Shi et al. 2002). However, handling errors of spatial data are somewhat different from the processing of conventional surveying data, as the sources of spatial data are diverse and complex. In addition, operations with spatial data are also complex, and are different from surveying adjustment methods, among which there are strict geometric conditions. Therefore, as well as traditional probability and statistics theory, other theoretical supports are required according to the intrinsic features of spatial data (Burrough and Frank 1996; Burrough et al. 1997; Fisher 1999; Goodchild and Jeansoulin 1998). Fuzzy set theory provides an important approach to dealing with spatial data, and has sporadically been adopted in the field of GIS (Cheng et al. 2001; Fisher 2000).

The connection between fuzzy set theory and spatial data quality control needs to explore the relationship between spatial data error and uncertainty theory. The uncertainty is the deficiency in the degree of knowledge about surveying data, and also the degree of unlikelihood and doubt about the validity of the survey results. Standard deviations, or multiples of these, are always used to express uncertainty. The surveying error is the difference between the measured value and the true value, and this is caused by imperfect processing of the survey data or unsatisfactory surveying conditions. As the true value is generally unknown, the actual value of the error is difficult to find. In theory, surveying adjustment methods allow surveying data to be closer to the true value. In the international "Guide to the Expression of Uncertainty in Measurement" (International Organization for Standardization 1995), it was stressed that surveying error and uncertainty are deemed to be two different concepts which should not be confused. Surveying error is really different from uncertainty by definition. However, both of them are used to express the relationship between the survey data and the true value. They play the same role in describing the confidence level of the survey data. In other words, surveying error is one of the descriptive methods of indicating surveying uncertainty. Goodchild (1999) argued that uncertainty is the difference between the true value and its expression in GIS. If the true value can be determined, error and precision are used to describe the uncertainty (Goodchild 1999). However, the true value is generally unknown, especially when it is related to the cognition of humans.

Randomness and fuzziness are the two conditions that result in the uncertainty of spatial data (Burrough and Heuvelink 1992). Randomness is the uncertainty of cognition generated by the inadequacy of the observation conditions. Fuzziness refers to the uncertainty of differentiation caused by the intermediary transitivity of objective differences (Burrough and Frank 1996; Fisher 2000; Zadeh 1965).

This type of intermediary transitivity may be due to the fuzziness of the ruler used to describe objective things, or the inevitability of unclear cognition about those things. GIS can model the real world. The model is based on the cognition and abstraction of the real world, and is the approximate reflection of the real world. The randomness and fuzziness of reality cognition bring uncertainty about the spatial data into GIS, and thus affect the spatial data quality and the results of spatial analysis.

The uncertainty of spatial data derived from the concept of fuzziness can be separated into several aspects, as follows.

1. Uncertainty comes from the vagueness of the distribution of geographical phenomena or the concept of a geographical entity

   Variation and fuzziness are two intrinsic attributes in nature which affect the accuracy of spatial data representation. For instance, the range of grassland is not always determinate; somewhere grassland always moves gradually toward forest or desert areas, or else there exists a transition area which reflects a smooth transition state from grassland to forest or desert. The boundary of soil units and the classification of vegetation type are usually fuzzy, and different operators often draw different classification maps, and so forth. If such fuzzy information does not undergo appropriate processing, the data imported into GIS will definitely have fuzzy characteristics, and result in uncertainty.

   This kind of fuzziness can be reflected in both graphics and the attribute of accuracy in the quality of the contents of spatial data. The intuitionistic reflection in the accuracy of graphics is an error. However, the real position of a boundary is difficult to determine owing to the vagueness of the concept. Therefore, the fuzziness of a concept may be reflected in the accuracy of the attribute. For example, the vagueness of vegetation classifications will result in mistakes when describing parcels of vegetation.
2. Uncertainty derived from spatial relationships

   In a qualitative description of a spatial relationship, there ubiquitously exist inaccurate terms. For example: what is "*nearby*" the village; land to the "*south*" (or "*north*") of the river is suitable for arable farming. In buffer analysis, the proximity to a river is usually described in terms like "in the area "*about*" 5 km to the "*north*" (or "*south*") of the river"; in visual interpretations of images, a ground feature may be described as residential houses, factory buildings, or other type of architecture. There are often inaccuracies when describing geographical properties. For instance, the descriptions of boundaries always are fuzzy, as multiple feature boundary lines often coincide with each other. The mixed pixels generated in the processing of remote sensing data and the overlapping in mode identification are also fuzzy. These emerge in the process of finding specified descriptions of geographical phenomena.

   This kind of vagueness in specific descriptions of spatial relationships is mainly reflected in attribute accuracy, logistic consistency, and the integrity and temporal accuracy of the quality and contents of spatial data. If the description of geographic phenomena is not clear, the accuracy of the attribute will always be affected, as well as the logistic consistency and integrity of the data. During a set time interval, spatial data may fail to accurately reflect the situation at that time. Therefore, the temporal accuracy will also be influenced.

3. Uncertainty derived from spatial analysis and spatial reasoning operations

   Spatial analysis is a reasoning process about knowledge, the results of which can provide spatial decision-making support to users. The language of human beings is similar to fuzzy semantic expressions, for example, we want to find the "*largest*" area affected when a reservoir bursts, or to consider the villages which are at a distance of "*about*" 200 m from the reservoir, and so forth. Sometimes we need to consider whether an area is "*suitable*" for a certain kind of crop. Here, "*suitable*" can be divided into fuzzy terms like "*totally unsuitable*," "*not very suitable*," "*suitable*," "*comparatively suitable*," and "*totally suitable*."

   Uncertainty generally exists in spatial data, and originates from the vagueness of the concept of spatial entities and spatial relationships. This uncertainty influences the quality of spatial data, and thus affects the results of GIS applications. Traditional Boolean set theory can deal with spatial entities with determinate boundaries and concepts in cognition. However, for those spatial entities with fuzzy boundaries and concepts, Boolean set theory fails to reflect the vagueness among them. In traditional approaches, such spatial entities would be modeled approximately, and accordingly, the approximate model would result in a loss of information. Such uncertainty needs fuzzy set theory.

4.2 Fuzzy Set Theory

Fuzzy set theory was first presented in 1965 by the famous cybernetics expert L.A. Zadeh in his ground-breaking paper *Fuzzy Sets* (Zadeh 1965). In his research on human thinking and judgment of the modeling process, he built up a theoretical system using rigorous mathematical methods to describe fuzzy phenomena. Fuzzy set theory is an extension of the traditional classic set theory. The aim of the extension is to overcome the accurate "either–or" bi-value logic of classic set theory. Thus, there is a smooth transition between elements and non-elements of a set, so that one element can partially belong to a set, but not completely belong or completely not belong to the set. The difference between a fuzzy set and a classic set is that the fuzzy set has explicitly put forward the terms of a membership function through which the degree of each element belonging to a set can be calculated. Set operations like intersection and union in classic set theory are still applicable in fuzzy sets.

4.2.1 Fuzzy Set

When people consider a specific problem, they always confine the issue within a limited range, which is the so-called universe, and is usually represented by capital letter U, V. The components in the universe are elements which are usually embodied by lowercase x, y. Given a universe U, a group of different elements in the universe is called a set, which is usually represented by A, B and so on. In classic set theory, the relationship of an element x with a set A has only two cases: $x \in A$ or $x \notin A$. However, the existence of vagueness in the objective world makes it impossible for the "either–or" thought in classical set theory to present all the relationships of each element within the set.

In classic set theory, an eigenfunction is used to depict the relationship between elements and a set. Each set A has an eigenfunction $C_A(x)$. If $x \in A$, then $C_A(x) = 1$; if $x \notin A$, then $C_A(x) = 0$.

$$C_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{4.1}$$

Eigenfunction $C_A(x)$ is a mapping from the universe U to a range [0, 1]. Usually it can separate elements in the set A from those outside of the set A. $C_A(x)$ is a binary value function which can only distinguish two situations, to be or not to be, and is applicable to objects with determinate definition. As it cannot distinguish the degree of membership, it is not suitable for fuzzy phenomena.

The basic idea of a fuzzy set makes the absolute affiliation relations in a classic set flexible. In the form of an eigenfunction, the grade of membership is not confined to 0 or 1, but can be any value between 0 and 1. Given a universe U and a membership function, each element x in U can be connected with a value $\mu_A(x)$ in [0, 1]. $\mu_A(x)$ is used to express the grade of membership of element x belonging to the set A. Here, A is a fuzzy set, and $\mu_A(x)$ is equivalent to eigenfunction $C_A(x)$. Its value is no longer confined to 0 and 1, and has expanded to any value between [0, 1].

Definition 4.1 Given a universe U and its mapping μ_A in the closed interval [0, 1]

$$\mu_A : U \to [0,1]$$

$$x \to \mu_A(x), x \in U$$

A fuzzy subset A in the universe U can be determined, and generally be referred to as a fuzzy set. $\mu_A(x)$ is called the grade of membership belonging to the fuzzy set A.

4.2.2 Fuzzy Set Operations

Operations between two fuzzy sets actually operate on the grade of membership point by point.

1. $\supseteq$ indicates inclusion

   Given A, B as two fuzzy sets in the universe U, if there are $\mu_A(x) \le \mu_B(x)$ for any $x \in U$, then B includes A, denoted $B \supseteq A$.

If $\mu_A(x) = \mu_B(x)$, fuzzy set A equals B, denoted A=B.

A fuzzy set with all membership at 0 is called a null set or an empty set, denoted Φ.

2. A^c indicates the complementary set of fuzzy set A

Given that A is a fuzzy set in the universe U, the complementary set A^c can be defined as follows:

$$\mu_{A^c}(x) = 1 - \mu_A(x)$$

3. $A \bigcup B$ indicates the union set of fuzzy sets A and B

Given two fuzzy sets A, B in a universe U, a new fuzzy set C is the union set of A and B. For any $x \in U$, the membership of x included by C can be determined by the larger of $\mu_A(x)$ and $\mu_B(x)$

$$C = A \bigcup B \Leftrightarrow \forall x \in U$$

$$\mu_C(x) = \max(\mu_A(x), \mu_B(x))$$

4. $A \bigcap B$ indicates the intersection of sets A and B, and can be defined as:

$$D = A \bigcap B \Leftrightarrow \forall x \in U$$

$$\mu_D(x) = \min(\mu_A(x), \mu_B(x))$$

5. Cut the operation of fuzzy set A

Given that A is a fuzzy set in a universe U, for any real number $\lambda \in [0,1]$, the λ -level cut set of fuzzy set A is

$$A_\lambda = \{x \mid \mu_A(x) \geq \lambda, x \in U\}$$

A_λ is a classic set. A fuzzy set is converted to a common set by the cut operation.

4.2.3 Fuzzy Relationships

There are various relationships in the world. The relationship between two objects is usually represented by a trenchant subset, such as terms like *x equals y* or *x is larger than y*, and so on.

Definition 4.2 Element x in a set A and element y in a set B can form an ordered pair (x, y). All these pairs (x, y) constitute a set which is a direct product of A and B, denoted $\mathrm{A} \times \mathrm{B}$.

$$\mathrm{A} \times \mathrm{B} = \{(x, y) \mid x \in A, y \in B\}$$

Definition 4.3 As for sets A and B, any subset R of their direct product $U \times V$ is called a binary relation between A and B, or simply referred to as a relation.

Given that both A and B are finite sets, the relation R can be signified as

$$R = \{r_{ij}\}_{m \times n}$$

Here, m stands for the number of elements in set A, n is the number of elements in set B, and $r_{ij} \in [0,1]$, $i = 1,2,\ldots,m; j = 1,2,\ldots,n$.

If R is a fuzzy set, it depicts the fuzzy relation between A and B. The value of elements of R can be defined as

$$r_{ij} = \mu_R(a_i, b_j)$$

where $\mu_R(a_i, b_j)$ stands for the grade of membership in the universe of $A \times B$.

4.2.4 Defining the Membership Functions

The membership function of a fuzzy set, usually expressed as $f_A(x)$, defines how the grade of membership of x in A is determined. There are two possible ways of deriving these membership functions (Metternicht 1999). The first approach, called the similarity relation model, resembles cluster analysis and numerical taxonomy in that the value of the membership function is a function of the classifier used (Robinson 1988). A common version of this model is the fuzzy *k*-means or *c*-means method, which is used for soil grouping, remote sensing image classification of cloud cover, and vegetation analysis (McBratney and de Gruijter 1992; Wang 1990). The second approach, known as the semantic import model, uses an a priori membership function with which individuals can be assigned a membership grade. This model is useful in situations where users have a good qualitative idea of how to group data, i.e., the exact associations of the standard Boolean model (Burrough 1989).

4.3 Applications in Previous Studies

According to the analysis above, fuzziness in spatial data can be divided into three aspects: the distribution of geographical phenomena or the concept of geographical entity, spatial relationships, and spatial analysis and spatial reasoning operations. Since fuzzy set theory was introduced into GIS, many scholars have made efforts to clarify the problems in each of these three aspects.

4.3.1 Fuzzy Representation of Geographical Entities and Their Distribution

Generally speaking, there is an implicit assumption in geographical entity modeling that the scope or boundary of spatial phenomena or entities can be defined accurately.

In a vector structure, the geometric shape of a geographical entity is represented by a point, line, or polygon which can be described accurately. The values of the attributes of such geographical entities are constant within the whole space range, such as land parcels, houses, roads, etc. However, the traditional modeling method is not appropriate when dealing with geographical entities with fuzziness in their definition and geographical distribution (called fuzzy objects) (Du et al. 2007; Schneider 1999). Fuzzy modeling can properly express fuzzy geographical objects caused by the vagueness of their geographical distribution or fuzzy definition, including natural, social, and cultural phenomena with consecutive change attributes. Schneider (1999) proposed accurate definitions for a fuzzy point, fuzzy line, and fuzzy polygon based on vector structure.

A fuzzy object is closely related to a field-based model (Du et al. 2007; Zhao et al. 2005). There are two types of field-based models: the numeric type and the category type. The numeric type of field-based model is suitable for modeling geographical phenomena whose attribute values change consecutively with location, such as topography fluctuations, gradual permeations from grassland to desert, etc. A function can be constructed to denote such an attribute value at any position. This model is a kind of numerical value of expression, and is accurate. In digital representations, field-based data models can be represented as the following continuous two-order relationship on a 2-D plane N^2:

$$R = \int_{(x,y)} \frac{\mu_R(x,y)}{(x,y)} \quad x, y \in N^2 \tag{4.2}$$

where fuzzy membership value $\mu_R(x,y)$ represents the attribute density of a surface feature character at point (x,y). That is to say, it stands for the extent to which a point belongs to one class (object). If $\mu_R(x,y)$ equals any one of both numbers {0, 1}, all the objects in real-life have crisp boundaries. If $\mu_R(x,y)$ is a numerical value in the interval [0, 1], R becomes a fuzzy set and the model can represent fuzzy geographical phenomena. This relationship can be expressed with a 2D matrix in which the row and column numbers are the coordinates of the spatial surface feature. For example, an urban area can be represented as shown in Fig. 4.1.

The values of the cells stand for the extent to which the cell belongs to the urban area. 1.0 indicates that the cell belongs entirely to the classification of urban area; 0 < the value < 1.0 means that the cell partly belongs to the urban area; 0, the cell cannot be characterized as urban area at all. The two-order relationship reflects the field view of geographic phenomena.

In the category type of field-based model, each position belongs to different types of attribute. As the attribute types are often qualitative and discrete, the key to this model is the classification system. That is to say, each position is given an attribute type. Each pixel in this model belongs to just one category, and the degree of membership is 1. Therefore, it cannot describe partial membership of fuzzy phenomena.

The numeric type of field-based model is relatively more suitable for describing a fuzzy object. However, because of the deficiency of functions in existing GIS to process fuzzy data, its application is not possible. Clementini and Di Felice (1997) a

0	0	0	0	0	0.1	0.1	0.1	0	0	0	0
0	0	0	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0	0
0	0.1	0.2	0.2	0.3	0.4	0.4	0.4	0.3	0.3	0.2	0.1
0.1	0.2	0.4	0.5	0.5	0.7	0.7	0.6	0.5	0.5	0.3	0.2
0.1	0.3	0.4	0.6	0.7	0.9	0.9	0.8	0.7	0.6	0.4	0.3
0.3	0.4	0.6	0.7	0.9	1.0	1.0	1.0	0.8	0.7	0.5	0.3
0.3	0.5	0.7	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.7	0.5
0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6
0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6

b

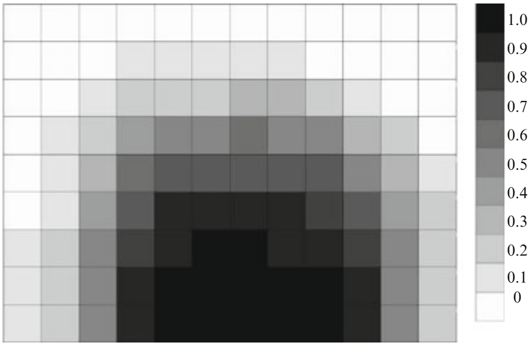


Fig. 4.1 Fuzzy representation of a spatial object—an urban area. (**a**) In numerical form. (**b**) In *gray* form. The hierarchy of *gray* values is shown on the *right* of (**b**)

put forward the concept of a broad boundary model. An object is formed by an interior, a broad boundary, and an exterior (Fig. 4.2a) (Clementini and Di Felice 1997). A broad boundary has a certain width and area, and is no longer a geometric line. The broad boundary model can usually be expressed as two areas: exterior and interior. The exterior area illustrates where the object may be located, and the interior area where the object must be located. The difference between the exterior area and the interior area is the broad boundary. This model uses a broad boundary to reflect the uncertainty of a fuzzy object. An object represented using the broad boundary model is called a broad boundary object. According to the complexity of the object, a broad boundary object can be defined as simple or complex. A simple broad boundary object is composed of a continuous interior, a continuous boundary, and a continuous exterior (see Fig. 4.2a), while a complex broad boundary object is a combination of several simple broad boundary objects. Cohn and Gotts (1996) advanced the "egg-yolk" approach to represent the uncertain area (Fig. 4.2b). The internal deep gray sub-region in the model is called the "egg-yolk", and the light gray sub-region outside is called the "egg-white" (Cohn and Gotts 1996). The "egg-white" stands for the uncertain part. The broad boundary model and the "egg-yolk" model can qualitatively describe the fuzzy extent, but cannot distinguish the membership of each pixel

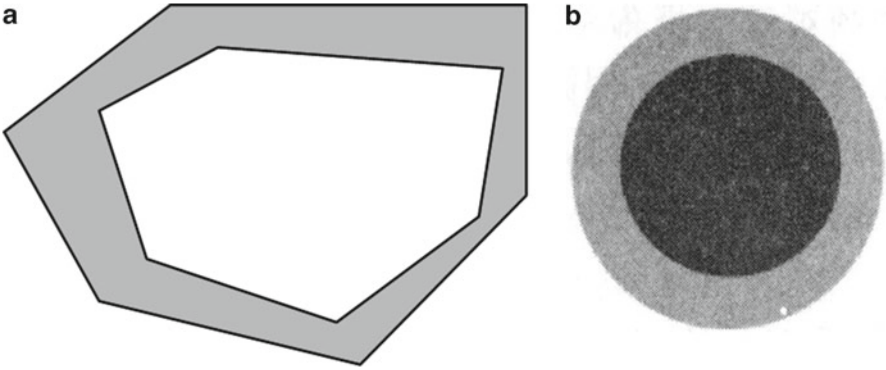


Fig. 4.2 Broad boundary region (**a**) and "egg-yolk" model (**b**) (Cohn and Gotts 1996)

belonging to the broad boundary or the "egg-white". A broad boundary region or "egg-white" can be obtained by a λ-level cut set for the fuzzy object (Fig. 4.2).

Du et al. (2007) summarized the source of a fuzzy object. There are three sources of fuzzy objects: the inherent model characteristics of the geographical phenomena (i.e., geographic distributions or geographic concepts are vague), the deficiency of spatial resolution (the geographic phenomena are accurate, but the spatial resolution is insufficient), and the derivation from existing fuzzy or non-fuzzy objects. The inherent fuzziness of geographical phenomena determines that a pixel does not completely belong to a certain category. There exists a certain transition or overlap area among categories. For the second source, as the spatial resolution in remote-sensing images is not high enough, fuzziness and hybrid pixels are generated. What the pixel represents on the ground is a synthesis of different adjacent objects. The third source comes from fuzzy operations of fuzzy or non-fuzzy objects (Stefanakis et al. 1999). The fuzzy operations include fuzzy overlay analysis, fuzzy buffer analysis, and fuzzy focus operations, etc. The result of these operations is also a kind of fuzzy object, just an outcome of the logical or arithmetic operations on the original objects. This derived object is similar to a fuzzy object in attributes and processing method except for the sources and meaning. Therefore, it can also be processed as a fuzzy object.

Cheng et al. (2001) classified fuzzy objects into three categories, fuzzy–fuzzy (FF), fuzzy–crisp (FC), and crisp–fuzzy (CF), and pixels can be classified into fuzzy objects according to different criteria (Fig. 4.3).

The FF model represents objects with uncertain thematic attributes and spatial scope. It allows different objects to overlap with each other. The FC model describes objects with a certain thematic content but uncertain spatial scope. The CF model describes objects with a certain spatial scope but uncertain thematic content. The FC and CF models are suitable for describing fuzzy objects which are separated spatially. An FC object's boundary is fuzzy with a precise interior. FC objects can overlap with each other, but CF objects cannot. Therefore, the traditional accurate object (crisp–crisp) model just describes objects with a determinate spatial scope and attribute range. According to the fuzzy object classification, Cheng et al. (2001)

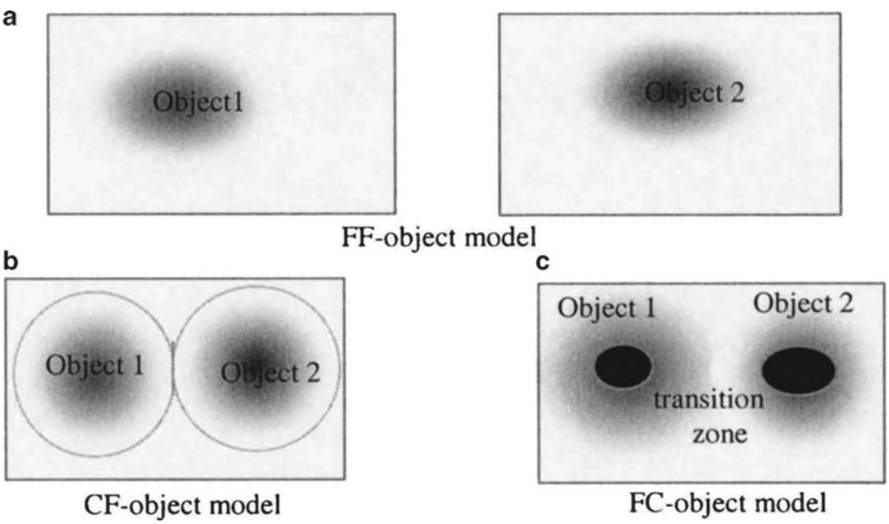


Fig. 4.3 Three fuzzy object models (**a**) FF-object model; (**b**) CF-object model; (**c**) FC-object model (Cheng et al. 2001)

defined different criteria using fuzzy and probability methods to extract fuzzy objects from the uncertainty classification results of remote sensing images.

4.3.2 Fuzzy Spatial Relationships

Spatial relationships may be caused by the geometric characteristics of spatial phenomena (the geographical position and shape of spatial phenomena) such as distance, direction, and connectivity, etc., or by the geometric and non-geometric characteristics of spatial phenomena together (including measurement attributes such as elevation value, slope values, etc., and the name attribute such as place names, etc.). For instance, the statistical correlation of spatially distributed phenomena, spatial autocorrelation, spatial interaction, spatial dependence, etc., belong to this kind of spatial relationship (Du et al. 2007). In qualitative spatial reasoning, it is common to consider the main spatial aspects of topology, direction, and distance, and to develop a system of qualitative relationships between spatial entities which cover this spatial aspect to some degree, and which appear to be useful from an application or cognitive perspective (Renz 2002). Therefore, this chapter chiefly focuses on the description of fuzzy spatial relationships such as topology, direction, and distance.

4.3.2.1 Fuzzy Description of Spatial Topology Relationships

A topological relationship refers to the property that remains the same in the process of topological transformation, such as translation, rotation, and scaling transformation, etc.

Topological relationships have always been the main content in spatial relationship research, and also an important component in spatial database queries and retrieval language. The 4-intersection model and the 9-intersection model are commonly used and accurate methods of describing topological relationships (Egenhofer and Franzosa 1991; Egenhofer and Herring 1991), and have received wide use and recognition in theoretical research and applications of GIS. The 9-intersection model can distinguish between 8 types of meaningful polygon–polygon topological relationships, 19 types of line–polygon topological relationships, and 33 types of line–line topological relationships. However, one shortcoming of the 4-intersection model and the 9-intersection model is that they can only describe topological relationships between determinate objects, while failing to describe the topological relationships of fuzzy objects.

Clementini and Di Felice (1997) replaced the mathematical boundary in the 9-intersection model with a broad boundary. The 9-intersection model derived from a combination of the interior, broad boundary, and exterior of two broad boundary objects is extended to describe fuzzy topological relationships. This is called the extended 9-intersection model. It can describe 44 topological relationships between simple broad boundary objects, and 56 topological relationships between complex broad boundary objects (Clementini and Di Felice 1997). Cohn and Gotts (1996) also proposed 46 types of topological relationships among fuzzy polygons base on the "egg-yolk" method.

In fact, a topological relationship can be formally described with a quintuple S_Topologic (*U*, *V*, *F*, *H*, *C*) (Du et al. 2007). *U* is an object set, *V* is a conceptual set for the topological relationship, *F* is a function mapping set, *H* is a partition function set of object space, and *C* is the range of values of *F* and *H*. For any topological relationship concept $v_i \in V$, a corresponding mapping $f_i \in F$ always exists with it. The function $f_i : U \times U \to C$ represents the consistency between topological relationships and the conceptual meaning of a topological relationship for any object *A* and *B* in the set *U*. The description of a topological relationship is implemented by mapping the topological relationship of *A* and *B* through the function f_i to set *V*. Thus, topological relationships of any two objects in *U* can be described by the concept in set *V*. For instance, when a polygon–polygon topological relationship is described in the 9-intersection model, *V* = {*disjoint*, *meet*, *overlap*, *cover*, *covered-by*, *contain*, *inside*, *equal*}, each concept in *V* corresponds to a matrix. *F* is a binary logic function set used to determine which concept matrix in set *V* is the same as the topological relationships of objects *A* and *B*. The value range of C is {0, 1}. Set *H* of a partition function set is the definition of the interior point, exterior point, and boundary point in point set topology (Gaal 1964). This definition is determinate, and its range of values is {0, 1}.

For fuzzy topological relationships, set *V* is identical to the 9-intersection model. The difference between fuzzy topological relationships and classic ones exists in *F*, *H*, and *C*. Three functions $h_1(x, y)$, $h_2(x, y)$, $h_3(x, y)$ in *H* define the membership of point (*x*, *y*) belonging to the interior, exterior, and boundary respectively, and each range of values is extended to [0, 1]. The value range of *C* in the function *F* $f_i : U \times U \to C$ is [0, 1], and the function $f_i : U \times U \to C$ is used to determine the

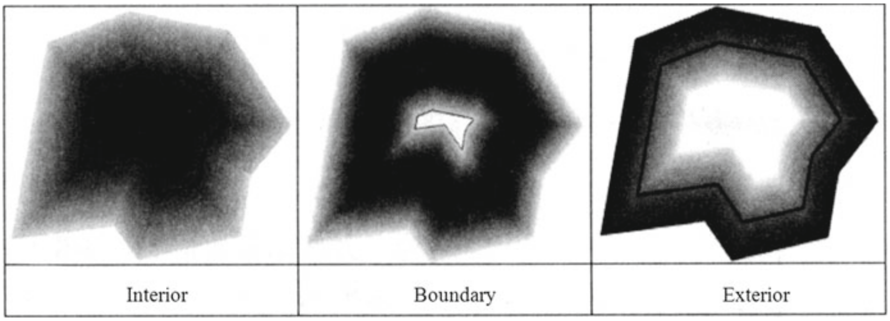


Fig. 4.4 Fuzzy partition of a polygonal object in a topological universe (Du et al. 2007)

topological relationships of objects *A* and *B* and the membership of each concept in set *V*.

In the case of polygonal objects, the definitions of interior point, exterior point, and boundary point in point set topology can be used to divide fuzzy space into three fuzzy sets as a boundary region, an interior region, and an exterior region (Du et al. 2007) (Fig. 4.4). As influenced by the uncertainty of spatial data, the boundary of a polygonal object is not simply referred to as its boundary with coordinates arrayed in a vector structure or a sequence of grids after the rasterization of the boundary in the raster structure. The boundary is extended to become a region spreading inward and outward to the object. The farther away from the boundary of an object a pixel is, the smaller the degree of membership of belonging to the object boundary becomes. The maximum membership degree is 1.0 at the polygon boundary.

4.3.2.2 Fuzzy Representation of Spatial Direction

Directional—also called orientational—relationships of spatial entities with respect to other spatial entities is usually given in terms of a qualitative category such as "*to the north of*" rather than using a numerical expression such as "*12 degrees*" (which is certainly more common in technical communications such as aviation). These are important and common-sense linguistic and qualitative properties used in everyday situations and qualitative spatial reasoning (Frank 1996). The direction of spatial entities is a ternary relationship depending on the located object, the reference object, and the frame of reference, which can be specified either by a third object or by a given direction. In the literature, one distinguishes between three different kinds of frames of reference, extrinsic ("external factors impose a direction on the reference object"), intrinsic ("the direction is given by some inherent property of the reference object"), and deictic ("the direction is imposed by the point of view from which the reference object is seen") (Hernàndez 1994). Given the frame of reference, directions can be expressed in terms of binary relationships with respect to the frame.

Most approaches to dealing with direction qualitatively are based on points as the basic spatial entities and consider only two-dimensional space. Frank (1991)

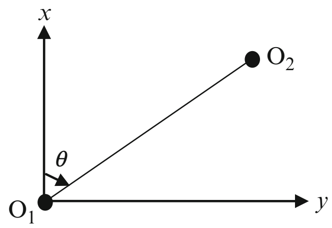


Fig. 4.5 Definition of azimuth θ from object O_1 to O_2

suggested different methods for describing the cardinal direction of a point with respect to a reference point in a geographic space, i.e., directions are in the form of [north, east, south, west] depending on the granularity (Frank 1991). Zhao et al. (2005) proposed a spatial direction model based on trigonometric functions. In this model, all objects are considered as a point—even those with irregular shape and size—and follow Frank's suggestion about directions (a centroid-based method, where the direction between two objects is determined by the angle between their centroids) (Fig. 4.5). The azimuth θ from object O_1 to object O_2 is computed. This angle, denoted by $\theta(O_1, O_2)$, takes values in $[0, 2\pi]$, which constitutes the universe on which primitive directional relations are defined. $Sin^2(\theta)$ and $cos^2(\theta)$ are chosen as fuzzy membership functions to describe the direction [north, east, south, west] with reference to the relative position relation functions proposed by Miyajima and Ralescu (1994) (Fig. 4.6). Miyajima and Ralescu (1994) used the square trigonometric function to illustrate the relative position relations [above, right, below, left] of segmented images. Square trigonometric functions are also suitable for directions in the form of [north, east, south, west] (Miyajima and Ralescu 1994). For instance, in Figure 4.5, if $\theta = 50°$, then the direction relationship is [0.4132, 0.5868, 0, 0] in the form of [north, east, south, west] according to Eqs. (4.3)–(4.6). This means that object O_2 is located to the north of object O_1 with 0.4132 of membership degree, and to the east with 0.5868 of membership degree. That is, $\mu_{\text{north}}(O_1, O_2) = 0.4132$, $\mu_{\text{east}}(O_1, O_2) = 0.5868$, $\mu_{\text{south}}(O_1, O_2) = 0$, $\mu_{\text{west}}(O_1, O_2) = 0$, and $\mu_{\text{north}}(O_1, O_2) + \mu_{\text{east}}(O_1, O_2) + \mu_{\text{south}}(O_1, O_2) + \mu_{\text{west}}(O_1, O_2) = 1$. Therefore, fuzzy membership functions not only show the characteristics of transition of the directional relationship, but also ensure the integrity of the definition of the direction for any target object.

$$\mu_{north}(\theta) = \begin{cases} \cos^2(\theta), & \frac{3\pi}{2} \le \theta \le 2\pi \; or \; 0 \le \theta \le \frac{\pi}{2} \\ 0, & \frac{\pi}{2} \le \theta \le \frac{3\pi}{2} \end{cases} \tag{4.3}$$

$$\mu_{east}(\theta) = \begin{cases} \sin^2(\theta), & 0 \le \theta \le \pi \\ 0, & \pi \le \theta \le 2\pi \end{cases} \tag{4.4}$$

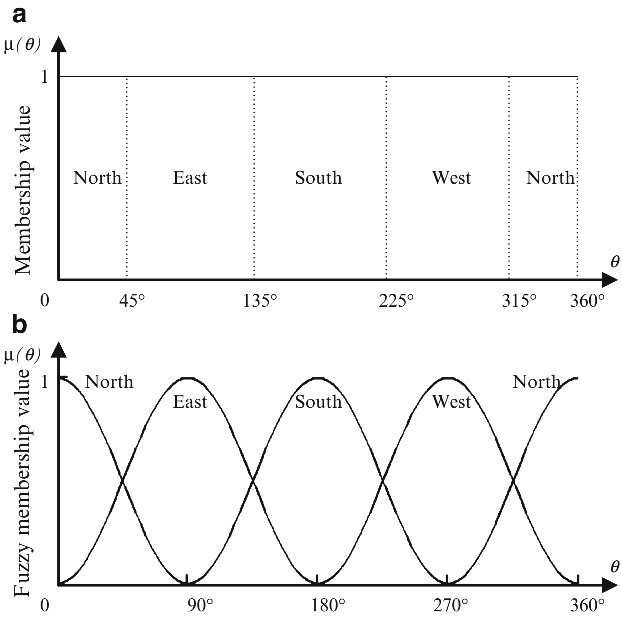


Fig. 4.6 An example of the classic (**a**) and fuzzy (**b**) classification of direction. θ stands for the azimuth from object O_1 to O_2

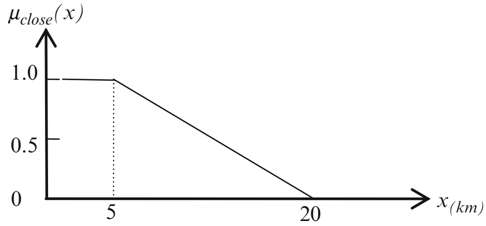
$$\mu_{south}(\theta) = \begin{cases} \cos^2(\theta), & \frac{\pi}{2} \le \theta \le \frac{3\pi}{2} \\ 0, & \frac{3\pi}{2} \le \theta \le 2\pi \; or \; 0 \le \theta \le \frac{\pi}{2} \end{cases} \tag{4.5}$$

$$\mu_{west}(\theta) = \begin{cases} \sin^2(\theta), & \pi \le \theta \le 2\pi \\ 0, & 0 \le \theta \le \pi \end{cases} \tag{4.6}$$

4.3.2.3 Fuzzy Description of Spatial Distance

In a spatial decision-making process, the distance relation between spatial entities always plays a key role. Dealing with distance is an important cognitive ability in our everyday life (Renz 2002). When representing distance, we usually use qualitative categories such as "*A is close to B*" (binary constraint) or qualitative distance comparatives such as "*A is closer to B than to C*" (ternary constraint), but sometimes also numerical values such as "*A is about 20 m away from B*". One can distinguish between absolute distance relations (the distance between two spatial entities) and relative distance relations (the distance between two spatial entities as compared with the distance to a third entity) (Guesgen and Albrecht 2000; Renz 2002).

Fig. 4.7 Membership function of "close to a city"



The choice of which relation should be used depends on the application universe and the requirements posed by decision-makes. For two individual locations *A* and *B*, which in general are abstracted as points, the Euclidean distance is given by the formula

$$d(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \tag{4.7}$$

where (x_A, y_A) and (x_B, y_B) denote the coordinates of two locations *A* and *B*, respectively.

Qualitative absolute distance relations are obtained, for example, by dividing the real line of distance into several sectors such as "*very close*," "*close*," "*commensurate*," "*far*," and "*very far*" depending on the chosen level of granularity (Hernàndez et al. 1995). In practice, we usually use one of the sectors. For instance, Figure 4.7 represents the "*close*" degree from a point on a map to a city.

$$\mu_{close}(x) = \begin{cases} 1, & x \le 5 \\ (20 - x)/15 & 5 < x \le 20 \\ 0, & x > 20 \end{cases} \tag{4.8}$$

where *x* denotes the distance (in kilometers) from the location to the city. The division values such as 5 km and 20 km are designed arbitrarily by decision-makers according to the understanding of their definition of a "*close*" degree.

4.3.3 Fuzzy Operations on Spatial Reasoning

Spatial reasoning is an approach for reasoning out unknown spatial relationships based on the determinate spatial relationships of objects. The reasoning is implemented through a symbolic operation based on implicit knowledge and rules. Spatial reasoning is a hot topic in fields such as GIS, artificial intelligence, and computer vision, and many spatial reasoning methods have been put forward. Du et al. (2007) systematically summarized the classification systems of spatial reasoning methods. Existing spatial reasoning methods are mainly concentrated in a single spatial relationship (Du et al. 2007) such as topological relationship reasoning by topological relationships, direction relationship reasoning by direction relationships, and so on. Fewer combinatorial spatial reasoning methods were presented. In addition, most of these spatial reasoning methods were developed for determinate geographical objects or spatial relationships, and sometimes it is difficult to adopt these methods for spatial relationships with uncertainty. Therefore, these methods need an extension and supplement for fuzzy geographic objects or fuzzy spatial relationships.

Hong et al. (1995) have researched combinatorial spatial reasoning methods for direction and distance relationships. The distance and direction between *A* and *C* can be reasoned according to the distance and direction between *A* and *B* and those between *B* and *C* (Hong et al. 1995). Sharma (1996) used a projection model to describe direction relationships, a 9-intersection model to describe topological relationships, and an interval model to describe qualitative distance relationships. Then single, combinatorial, and integrated spatial reasoning methods were proposed (Sharma 1996). As he used a projection model to describe a direction relationship between polygonal objects, this method cannot exactly express the actual directional relationship of such polygonal objects. In particular, because he focused on spatial reasoning for two single direction relationships rather than for single to multinomial direction relationships, multinomial to single direction relationships, and multinomial to multinomial direction relationships, the efficiency and accuracy of spatial reasoning results are limited.

A broad boundary model has been used to describe fuzzy objects (Clementini and Di Felice 1997; Worboys and Clementini 2001). Clementini and Di Felice (1997) proposed 44 types of topological relationships using the extended 9-intersection model based on the broad boundary model, and Cohn and Gotts (1996) put forward 46 types of topological relationships using the "egg-yolk" theory. As the number of topological relationships based on the broad boundary model or the "egg-yolk" theory goes beyond the range of cognition of a human being, such methods are inconvenient for topological relationship reasoning. Some scholars have also put forward advanced models to extend the fuzzy spatial reasoning method. Du et al. (2007) proposed a quadruple model to describe the topological relationship of objects with broad boundaries, and topological relationship reasoning was implemented based on the quadruple model.

Zhao et al. (2005) proposed a field-based integrated spatial reasoning model for the case of the constraint satisfaction problem (CSP). They argued that knowledge about spatial entities or about the relationships between spatial entities are often given in the form of constraints. Ordinarily, binary constraints such as "*the primary school should be laid out in the north of the residential area*," ternary constraints such as "*the primary school should be laid out between residential area A and residential area B*," or in general, *n*-ary constraints restrict the universe of 2, 3, or *n* variables. Problems like these are formalized as a constraint satisfaction problem: given a set of variables *R* over a universe *D* and a set *A* of constraints on the variables *R* (Renz 2002). CSP is a powerful general framework in which a variety of combinatorial problems can be expressed (Creignou et al. 2001; Marriott and Stuckey 1998). The aim of CSP is to assign values to the variables subject to specified constraints. In fact, it is the most popular reasoning

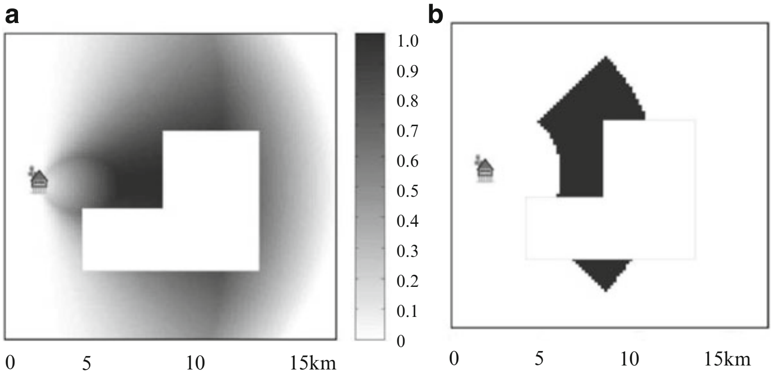


Fig. 4.8 The result of combinatorial spatial reasoning (**a**) and of traditional spatial reasoning (**b**)

method used in qualitative spatial reasoning (Renz 2002), and a common problem in spatial decision-making processes such as the examples described above.

Ladkin and Maddux (1994) formulated binary CSPs as relation algebras as developed by Tarski (1941). This allows binary CSPs to be treated in a uniform way. In a fuzzy domain, the relation algebras constitute fuzzy logic reasoning. Fuzzy logic reasoning, one of the application domains of fuzzy relationship generalization, is the fundamental basis of fuzzy spatial reasoning. It implements tasks through logical operations based on normal relation algebra theory. The operations can be extended to n sets of fuzzy relationships, i.e., the operation is applicable to multiple fuzzy sets. Assuming that there are n sets of fuzzy relationships, operations can be expressed uniformly as

$$\mu_R(z) = \bigotimes_{i=1}^{n} \mu_{A_i}(z) \quad z \in C \tag{4.9}$$

where ⊗ denotes operators *union*, *intersection*, and c*omplement,* respectively, and A_i stands for multiple fuzzy relationships.

Zhao et al. (2005) used this model to implement a combinatorial fuzzy reasoning including direction and distance relationships. A task in combinatorial fuzzy reasoning is to find a suitable location for a special factory given certain constraining factors.

(a) The factory must be located to the east of the environmental monitoring station.
(b) The factory must not be far from the environmental monitoring station.
(c) The factory must not be situated on land suitable for agriculture.

They compared the results of the combinatorial fuzzy reasoning model (Fig. 4.8a) with those of traditional spatial reasoning (Fig. 4.8b). It is evident that the information in the combinatorial fuzzy reasoning result is more abundant and more detailed than that in the result of the traditional approach. The combinatorial fuzzy reasoning model gives decision-makers more chances to choose a suitable result as it provides the degree of suitability to the proposition proposed by the users.

4.4 Conclusion and Future Prospects

Fuzziness is an inherent characteristic in nature and in the language of human beings. As GIS can model and digitize the real world, it needs to be able to deal with the fuzziness existing in the real world. In addition, the model of the real world in GIS is based on the cognition of humans, and the concept and methods of modeling and analysis embedded in GIS software are always influenced by human cognition. Spatial decision-making support provided by GIS should relate to human cognition. Accordingly, the representation and analysis of the real world in GIS should be geared to human language, so that the representation, analysis process, and results can completely satisfy the natural mode of expression of humans.

Scholars began to pay attention to the quality of the spatial data in GIS soon after the emergence of GIS in the 1960s, and they adopted fuzzy set theory to deal with spatial geographical phenomena and spatial analysis methods. In the early years, the application of fuzzy set theory in GIS emphasized the analysis of the sources of fuzzy phenomena in spatial data and spatial analysis methods. These sources include the fuzziness in three aspects of geographical distributions, or the concept of spatial entity, spatial relationships, and spatial reasoning. The exploration of sources indicated the research directions for the application of fuzzy set theory in GIS. After that, methods of representing fuzzy geographical objects in GIS were explored based on vector and raster structures. However, most of these representational methods can be deemed to be concept models, and do not deviate from traditional spatial data structure. In fact, field-based raster models or vector models of fuzzy geographical objects are always more complex than traditional data structure. Therefore, such models cannot be widely used for storing fuzzy spatial data. As spatial relationships and spatial reasoning are based on determinate geographical objects, research into them was also influenced by representational methods of spatial relationships and spatial reasoning. Researchers have not abandoned the struggle with fuzzy spatial data. The broad boundary model and the "egg-yolk" theory were proposed in order to extend traditional topological relationships, and the qualitative directional and distance relationships of fuzzy objects were also researched using fuzzy set theory and membership functions. In particular, some fuzzy reasoning methods have been put forward to solve fuzzy geographical problems.

These, however, are only the beginning of longer term collaborative research efforts. The application of fuzzy set theory in GIS is still developing. Both fuzzy mathematics itself and the concept of fuzzy geographical objects and analysis need innovation.

1. Further work will have to emphasize research into the measurement of fuzzy uncertainty. Based on this idea, a membership function model should be constructed for the fuzzy uncertainty of spatial data derived from different sources.
2. Based on geographical ontology, formalized methods of expression should be explored for fuzzy objects. In addition to traditional vector and raster structures, we may find a new integrative data structure in which fuzzy information can be stored and visualized conveniently in computer systems.

3. Further research should also focus on the formal definitions of combinatorial fuzzy spatial reasoning operations and predicates, with the integration of fuzzy spatial data types into query languages, and with aspects of implementation leading to sophisticated data structures for efficient algorithms for the operations.
4. The promulgation of fuzzy information in the process of spatial reasoning is a key point for users who pay attention to the results of spatial analysis. We will have to explore the law of promulgation and to assess its effect on the results of spatial analysis. In that way we could provide users with a confidence range of fuzzy spatial reasoning results.

References

Burrough PA (1989) Fuzzy mathematical methods for soil survey and land evaluation. J Soil Sci 40:477–492

Burrough PA, Frank AU (1996) Geographic objects with indeterminate boundaries. Taylor & Francis, London

Burrough PA, Heuvelink GBM (1992) The sensitivity of Boolean and continuous (Fuzzy) logical modeling to uncertain data In: Proceedings of 3rd European conference on geographical information systems, Munich, Germany, pp 1032–1041

Burrough PA, van Gaans PFM, Hootsmans R (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77:115–135

Cheng T, Molenaar M, Lin H (2001) Formalizing fuzzy objects from uncertain classification results. Int J Geogr Inf Sci 15:27–42

Clementini E, Di Felice P (1997) Approximate topological relations. Int J Approx Reason 16:173–204

Cohn A, Gotts N (1996) The "Egg-Yolk" representation of regions with indeterminate boundaries. In: Burrough P, Frank AU (eds) Geographic objects with indeterminate boundaries. Taylor and Francis, London, pp 171–187

Creignou N, Khanna S, Sudan M (2001) Complexity classifications of Boolean constraint satisfaction problems. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia

Doucette M, Paresi C (2000) Quality management in GDI. In: Groot R, MacLaughlin J (eds) Geospatial data infrastructure: concepts, cases, and good practice. Oxford University Press, Oxford, pp 85–96

Du S, Wang Q, Qin Q (2007) Representation and combined reasoning of spatial relationships. Science Press (in Chinese), Beijing

Egenhofer MJ, Franzosa R (1991) Point-set topological spatial relations. Int J Geogr Inf Syst 5:161–174

Egenhofer MJ, Herring J (1991) Categorizing binary topological relations between regions, lines and points in geographic databases. Technical Report, Department of Surveying Engineering, University of Maine

Fisher P (2000) Sorites paradox and vague geographies. Fuzzy Set Syst 113:7–18

Fisher PF (1999) Models of uncertainty in spatial data. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems: principles and technical issues, vol 1. John Wiley & Sons, New York, pp 191–205

Frank AU (1991). Qualitative spatial reasoning about cardinal directions. In: Proceedings of Auto-Carto 10, Baltimore, Maryland, pp 148–167

Frank AU (1996) Qualitative spatial reasoning: cardinal directions as an example. Int J Geogr Inf Syst 10:269–290

Gaal SA (1964) Point set topology. Academic, Duluth

Goodchild MF (1992) Geographical information science. Int J Geogr Inf Syst 6:31–46
Goodchild MF (1999) Future directions in geographic information science. Geogr Inform Sci 5:1–8
Goodchild MF, Dubuc O (1987) A model of error for choropleth maps with applications to geographic information systems. In: Proceedings of AUROCARTO8, Falls Church, pp 165–174
Goodchild MF, Gopal S (1989) Accuracy of spatial database. Taylor & Francis, London
Goodchild MF, Jeansoulin R (1998) Data quality in geographic information – from error to uncertainty. Hermes, Paris
Guesgen HW, Albrecht J (2000) Imprecise reasoning in geographic information systems. Fuzzy Sets Syst 113:121–131
Hernàndez D (1994) Qualitative representation of spatial knowledge. Sparger-Verlag, Berlin
Hernàndez D, Clementini E, Di Felice P (1995) Qualitative distances. Lect Notes Comput Sci 998:45–57
Heuvelink G, Burrough P, Stein A (1989) Propagation of errors in spatial modelling with GIS. Int J Geogr Inf Syst 3:303–322
Hong J, Egenhofer M, Frank AU (1995) On the robustness of qualitative distance and directions reasoning. In: The 12th international symposium on computer assisted cartography, Charlotte, North Carolina, pp 301–310
International Organization for Standardization (ISO) (1995) Guide to the expression of uncertainty in measurement. ISO, Geneva
Ladkin PB, Maddux R (1994) On binary constraint problems. J ACM 41:435–469
Marriott K, Stuckey P (1998) Programming with constraints: an introduction. MIT Press, Cambridge
McBratney A, de Gruijter J (1992) A continuum approach to soil classification by modified fuzzy k-means with extra grades. J Soil Sci 43:159–175
Metternicht G (1999) Change detection assessment using fuzzy sets and remotely sensed data: an application of topographic map revision. ISPRS J Photogramm Remote Sens 54:221–233
Miyajima K, Ralescu A (1994) Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations. Fuzzy Set Syst 65:225–236
Morrison JL (1995) Spatial data quality. In: Guptill SC, Morrison JL (eds) Elements of spatial data quality. Elsevier Science, Tokyo, pp 1–12
Mowrer TH, Congalton RG (2000) Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Taylor & Francis, London
Östman A (1997) The specification and evaluation of spatial data quality. In: Proceedings of the 18th ICA international cartographic conference. Swedish Cartographic Society, Sweden, pp 836–847
Renz J (2002) Qualitative spatial reasoning with topological information. Springer, Berlin
Robinson VB (1988) Some implications of fuzzy set theory applied to geographic databases. Comput Environ Urban Syst 12:89–97
Schneider M (1999) Uncertainty management for spatial data in database: fuzzy spatial data types. Lect Notes Comput Sci 1651:330–351
Sharma J (1996) Integrated spatial reasoning in geographic information systems: combining topology and direction. Ph.D. Dissertation, Department of Surveying Engineering, University of Maine
Shi W, Fisher P, Goodchild MF (2002) Spatial data quality. CRC Press, London
Stefanakis E, Vazirgiannis M, Sellis T (1999) Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. Int J Geogr Inf Sci 13:657–675
Tarski A (1941) On the calculus of relations. J Symb Log 6:73–89
Wang F (1990) Improving remote sensing image analysis through fuzzy information representation. Photogramm Eng Remote Sens 56:1163–1169
Worboys M, Clementini E (2001) Integration of imperfect spatial information. J Vis Lang Comput 12:61–80
Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
Zhao Y, Murayama Y, Zhang Y (2005) Field-based fuzzy spatial reasoning model for geographical information systems: case of constraint satisfaction problem. Theory Appl GIS 13:21–31

Chapter 5
Spatial Prominence and Spatial Weights Matrix in Geospatial Analysis

Changping Zhang

In spatial statistical analyzes of geographical phenomena, a region or city under study might be divided into some small areal units such as a regular square tessellation, or into irregular shaped administrative units which have different spatial characteristics. If we are using geographical information science (GIS) to support the analysis, irregular areal units such as *cho* in Japan are usually represented as one such polygon with geometric attributes. In spatial statistics, an areal unit that has special geometric attributes and maintains significant spatial correlation and spatial interaction close to adjacent units is called a prominent areal unit or an important areal unit. The prominence of areal units can be measured by a prominence or influence-centrality index, which is obtained by using eigenfunctions or the Markov chains method from a spatial weights matrix (Tinkler 1972; Griffith and Jones 1980; Boots 1982; Bavaud 1998; Zhang and Murayama 2003).

In this chapter, after a review of the method of creating a spatial weights matrix, we will consider some types of weight function definitions in order to create different measures of prominent areal units, and use them to analyze the urban spatial pattern in Matsudo City, Chiba Prefecture.

5.1 A Review of Creating a Spatial Weights Matrix

The spatial weights matrix formed by weight coefficients is an integral part of spatial modeling. It is defined as the formal expression of spatial dependence between observations (Anselin 1988). In thinking about a type of spatial weights

C. Zhang (✉)
Faculty of Regional Development Studies, Toyo University, Tokyo, Japan

Formerly in Graduate School of Geoscience, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: cp-zhang@toyo.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_5, © Springer Japan 2012

matrix ***W***, Getis (2009) indicated that at least three ways exist, i.e., theoretical, topological, and empirical points of view. According to a theoretical point of view, the ***W*** matrix is exogenous to any system, and should be based on a preconceived matrix structure. Usually this structure is based on a theory of distance decay, which may come from notions of spatial interaction and gravity models. The topological viewpoint arose from a need to depict the actual configuration of the areal units contained within a study region. For example, a long, narrow areal unit would be represented differently than a short, wide unit. In this case, the matrix ***W*** might be specified in different ways, e.g., number of neighbors, length of side, or proportion of perimeter in common. Cliff and Ord (1969) said, "With a flexible system of weights, the researcher can highlight those features of a study area which he believes to be important." The empirical viewpoint approach implies that spatial dependence can be detected in the variables under study. A local autocorrelation point of view should be used to identify the exact level of spatial autocorrelation surrounding any given observation, and then a ***W*** matrix can be created.

Since the 1960s, many researchers have attempted to create a proper dependence representation in spatial weights matrix ***W***. With data in a raster model, ***W*** was constructed in the rook's case or the queen's case definition of neighbors. Using vector data, two types of ***W*** could be used. The first is called a simple binary connectivity definition, and uses a discrete function where spatial entities are assumed to be adjacent if and only if they share a common boundary. The nature of the interaction of the spatial phenomenon under study cannot always be captured by a simple binary proximity measurement. In this case, based on theoretical and conceptual considerations, a generalized weight function can be used, as advanced by Cliff and Ord (1981), which offers flexibility in defining spatial proximity. Typical functions incorporate the distance between the geographical centroids of areal units and/or the length of a common boundary between areal units (Can 1996).

Getis and Aldstadt (2004) summarized previous attempts to create a spatial weights matrix and identified many different types of weighting schemes, such as:

1. Spatially contiguous neighbors
2. Inverse distances raised to some power
3. Length of shared borders divided by the perimeter
4. Bandwidth as the *n*th nearest neighbor distance
5. Ranked distances
6. Constrained weights for an observation equal to some constant
7. All centroids within distance *d*
8. *n* nearest neighbors, and so on

Some of the newer schemes include:

1. Bandwidth distance decay
2. Gaussian distance decline
3. "Tri-cube" distance decline function

Furthermore, Getis and Aldstadt (2004) also proposed a spatial weights matrix based on the G_i^* local statistic (Ord and Getis 1995), which accounts for the spatial association extant within any region that has been divided into its constituent parts. In a series of simulation experiments, the matrix was compared with well-known spatial weights matrix specifications—two different contiguity configurations, three different inverse distance formulations, and three semi-variance models—and performed best according to the AIC (Akaike information criterion) and the autocorrelation coefficient evaluation.

Zhang and Murayama (2000) suggested a concept and algorithm of k-order neighbors based on Delaunay's triangulated irregular networks (TIN), and redefined Getis and Ord's (1992) local spatial autocorrelation statistic as $G_i(k)$ with weight coefficient $w_{ij}(k)$ based on k-order neighbors for the study of local patterns in spatial attributes.

Although the choice of a spatial weights matrix specification for spatial statistical analysis is not clear-cut and seems to be governed primarily by convenience or convention, Griffith (1996) proposed some explicit guidelines on specifications of the spatial weights matrix, and concluded that relatively large numbers ($n > 60$) of areal units should be employed in a spatial statistical analysis, and low-order spatial models should be given preference over higher-order ones.

5.2 Spatial Weights Matrix Formed by Weight Coefficients

In this section, we define four types of weight coefficient and create the spatial weight matrices formed by these coefficients for a simple hypothetical region.

5.2.1 Binary Weight

If we try to study spatial interrelationships distributed over a set of areal units, the spatial structure of units might be defined as the spatial contiguity, which is treated as an $n \times n$ spatial weights matrix $\boldsymbol{W}$ with binary variables (Cliff and Ord 1981; Zhang and Murayama 2003). The binary weight is represented as follows:

$$w_{ij} = \begin{cases} 1 & \text{if areas } i \text{ and } j \text{ are contiguous;} \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

The binary spatial weights matrix $\boldsymbol{W}$ has a symmetric form. If $\boldsymbol{W}$ is row-standardized:

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j \in J} w_{ij}}$$

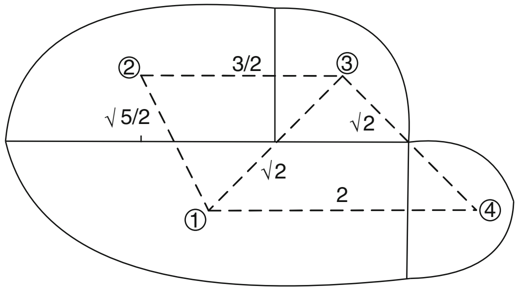


Fig. 5.1 A hypothetical region

where J is the set which includes all areal units that are contiguous with i, and w_{ij}^* satisfies non-negative $w_{ij}^* \geq 0$ and $\sum_{j \in J} w_{ij}^* = 1$ (for all $i = 1, \ldots, n$). The binary weight is often used in spatial autocorrelation analysis with regular and irregular areal units. A hypothetical example will clarify the use of this index. Imagine a region that is divided into four units (Fig. 5.1), units 1 and 2, 1 and 3, 1 and 4, and 2 and 3 are contiguous.

The binary spatial weights matrix and its standardized form in Fig. 5.1 are shown as follows.

$$W_b = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad W_b^* = \begin{pmatrix} 0 & 0.3333 & 0.3333 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

5.2.2 Distance Decay Weight

Tobler (1970) referred to the "first law of geography: everything is related to everything else, but near things are more related than distant things." In research into the interrelation and interaction, the phenomenon in which the contiguity or interaction decrease with increasing distance between two areal units is known as distance decay. The distance decay weight is usually expressed as a reciprocal of distance.

$$w_{ij} = \frac{1}{d_{ij}} \tag{5.2}$$

where d_{ij} is the distance from the center of unit i to the center of neighboring unit j. Generally, this distance can be a straight-line distance, a road distance, a time distance, a cost distance, a mental distance, or some other distance. The distance decay spatial weights matrix and its standardized form in Fig. 5.1 are shown below.

$$W_d = \begin{pmatrix} 0 & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{2}{\sqrt{5}} & 0 & \frac{2}{3} & \frac{2}{\sqrt{29}} \\ \frac{1}{\sqrt{2}} & \frac{2}{3} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{2}{\sqrt{29}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \qquad W_d^* = \begin{pmatrix} 0 & 0.4256 & 0.3365 & 0.2379 \\ 0.4628 & 0 & 0.3450 & 0.1923 \\ 0.3398 & 0.3204 & 0 & 0.3398 \\ 0.3168 & 0.2353 & 0.4480 & 0 \end{pmatrix}$$

5.2.3 Generalized Weight

A generalized weight definition is based on the distance and length of the boundary between areal units. In this case, the elements of spatial weights matrix $\boldsymbol{W}$ are

$$w_{ij} = \frac{l_{ij}}{d_{ij} \sum_{j \in J} l_{ij}} \quad (i = 1, \ldots, \mathrm{n}) \tag{5.3}$$

where J is the set which includes all units which are contiguous with unit $\boldsymbol{I}$, l_{ij} is the length of the join between units i and j. $\sum_{j \in J} l_{ij}$ is the length of all joins for unit i, i.e., its perimeter, and d_{ij} is the distance from the center of unit i to the center of neighboring unit j. $\frac{l_{ij}}{\sum_{j \in J} l_{ij}}$ is simply the proportion of the perimeter of unit i which is contiguous with unit j. This weighting function assumes that the interaction between two units will increase with an increase in the length of the common boundary, and will decrease with an increase in the distance between their geographical centers. The generalized spatial weights matrix and its standardized form in Fig. 5.1 are shown below.

$$W_g = \begin{pmatrix} 0 & \frac{1}{\sqrt{5}} & \frac{1}{4\sqrt{2}} & \frac{1}{8} \\ \frac{4}{3\sqrt{5}} & 0 & \frac{2}{9} & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{pmatrix} \qquad W_g^* = \begin{pmatrix} 0 & 0.5971 & 0.2360 & 0.1669 \\ 0.7285 & 0 & 0.2715 & 0 \\ 0.5148 & 0.4852 & 0 & 0 \\ 1.0000 & 0 & 0 & 0 \end{pmatrix}$$

5.2.4 k-Order Neighbors Weight

Zhang and Murayama (2000) proposed a concept of *k*-order neighbors based on Delaunay's triangulated irregular network (TIN), with centroids of areal units and a definition of *k*-order neighbors' weight. For example, Fig. 5.2b shows the result of drawing a TIN based on the point distribution in Fig. 5.2a. The *k*-order neighbors in the Delaunay triangulation network are defined as follows. Those points directly connected to a point *v* by the edges of the TIN are called the "nearest neighbors," or "first-order neighbors" in terms of this point. Then those points which are directly connected to the first-order neighbors of that point are not first-order neighbors themselves, but are called "second-order neighbors." By continuing this process, we can define *k*-order neighbors for any point, *v*.

The weight coefficient $w_{ij}(k)$, which measures the proximity of *k*-order neighbors in a point distribution, is determined as follows. $w_{ij}(k)$ is a binary coefficient which is set to 1 if point *j* is a *k*-order neighbor of point *i*, otherwise it is 0.

$$w_{ij} = \begin{cases} 1 & \text{Point } j \text{ is a } k\text{ - order neighbor of point } i; \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

The defined spatial weight matrix $\{w_{ij}(k)\}$ is a symmetrical binary matrix.

The first-order neighbor spatial weights matrix and its standardized form in Fig. 5.1 are shown below.

$$W_k = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad W_k^* = \begin{pmatrix} 0 & 0.3333 & 0.3333 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \\ 0.3333 & 0.3333 & 0 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

5.3 Prominence of Irregular Areal Units

5.3.1 Eigenfunctions Method

In some of the geographical literature, it is suggested that the principal eigenvector can be used to measure the prominence of irregular areal units (Tinkler 1972; Griffith and Jones 1980; Boots 1982). The eigenfunction of a spatial weights matrix is

$$\phi(\lambda) = \det\left(\lambda I - W\right) = \sum_{i=0}^{n} a_i \lambda^{n-i} = 0. \tag{5.5}$$

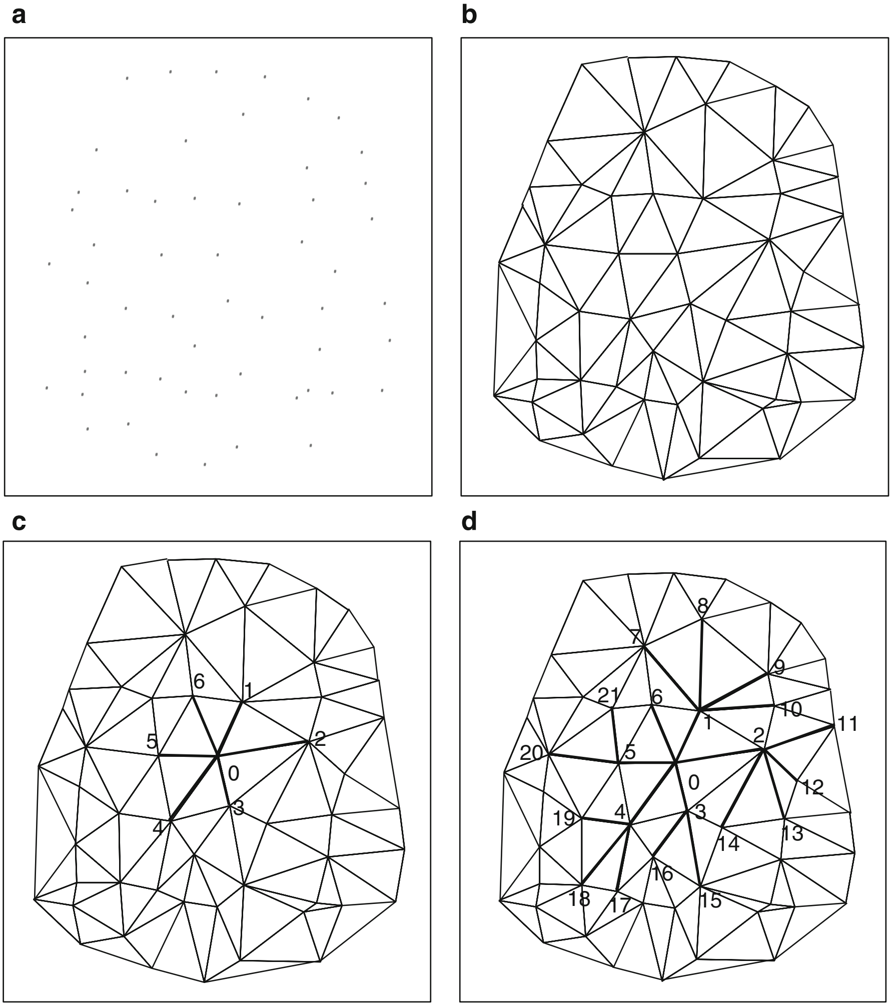


Fig. 5.2 *k*-order neighbors based on Delaunay triangulation. (**a**) Point distribution, (**b**) Delaunay triangulation, (**c**) nearest neighbors, (**d**) second-order neighbors (reproduced from Zhang and Murayama 2000)

The eigenvalues are the roots of $\boldsymbol{W}$. Since spatial weights matrix $\boldsymbol{W}$ is real, the λ_i is real and may be ordered as $\lambda_1 \geq \ldots \geq \lambda_n$. Corresponding to each λ_i is an eigenvector v_i, which satisfies

$$\boldsymbol{W}v_i = \lambda_i v_i \quad (i = 1, 2, \ldots, n). \tag{5.6}$$

The elements of v_i corresponding with the largest eigenvalue λ_i provide a measure of the relative position of each areal unit since the magnitude of the element is related to the centrality or prominence. Generally, a unit located in a central region of a city possesses a larger element of v_i, and a unit with a smaller element might be located near the city limits.

5.3.2 Markov Chain Method

The standardized spatial weights matrix $\boldsymbol{W}^*$ with elements of row-standardized variable w_{ij}^* is identical to the Markov chain transition matrix. Assuming the chain is ergodic ($(w^n)_{ij} > 0$ for some n (>0) and all i, j), we can obtain a unique stationary distribution $p_j > 0$, $(\sum_j p_j = 1.0)$ with the solution

$$W^T p = p. \tag{5.7}$$

or

$$\sum_j p_j w_{ij} = p_j$$

The vector $\boldsymbol{p}$, $p^T = (p_{1,} \cdots p_n)$, is an eigenvector of W^T with a corresponding eigenvalue of 1, whereas w_{ij} is a measure of the relative influence of unit j on unit i, and p_i can be interpreted as the total influence of unit i on the total region of a city. p_i will be further referred to as a prominence index (Bavaud 1998). We also call this index spatial prominence p_i here because its spatial attributes are necessary to calculate it.

5.3.3 Examples

As mentioned above, different prominences can be obtained from different spatial weight matrices. We use the standardized spatial weight matrices to calculate the prominence of the units in Fig. 5.1. The elements of eigenvector $\boldsymbol{p}$ of the binary spatial weights matrix $\boldsymbol{W}_b^*$ are

$$p_1 = 0.375, p_2 = 0.250, p_3 = 0.250, p_4 = 0.125.$$

Areal unit 1 has the largest value of prominence because it is contiguous to three other units. The prominence of unit 2 is the same value as that of unit 3 because they are all contiguous to two units. As unit 4 is only contiguous to unit 1, it has the least prominence.

The elements of eigenvector $\boldsymbol{p}$ of distance decay $\boldsymbol{W}_d^*$ are

$$p_1 = 0.273, p_2 = 0.251, p_3 = 0.270, p_4 = 0.205.$$

The value of the prominence of unit 1 is similar to that of unit 3, because the distances which link them to other units are almost equivalent. The prominences of units 2 and 4 are both smaller than those of units 1 and 3 since they are far apart from each other. As unit 2 is nearer to unit 1 than to unit 4, its prominence is larger than that of unit 4.

The elements of eigenvector $\boldsymbol{p}$ of generated $\boldsymbol{W}_g^*$ are

$$p_1 = 0.409, p_2 = 0.335, p_3 = 0.188, p_4 = 0.068.$$

Areal unit 1 has the largest prominence value because it possesses the largest unit and the longest common boundary with other units. The prominence of unit 2 is larger than that of unit 3 since unit 2 has a longer boundary and a shorter distance joined to unit 1 than unit 3. The prominence of unit 4 is smallest of the four 4 units.

The elements of $\boldsymbol{p}$ of first-order neighbors $\boldsymbol{W}_k^*$ are

$$p_1 = 0.3, p_2 = 0.2, p_3 = 0.3, p_4 = 0.2.$$

As units 1 and 3 are both directly connected to three other units by the edges of the TIN, their prominences are all 0.3. Similarly, as units 2 and 4 are both connected to only two other units, their prominences are smaller than those of units 1 and 3, and are equal to 0.2. Therefore, the variation of prominence based on the weight of first-order neighbors is relatively smaller and its sensitivity is lower compared with other types of weight.

The variation in the prominence of areal units depends on the definitions of the spatial weights matrix. The binary weight is affected by the topological attributes, and the generated weight is affected by the geometric attributes. The distance decay is completely determined by the distance between two units. However, the influence of geometric attributes is not clearly reflected in the weight of first-order neighbors.

5.4 Application

Matsudo is a medium-size city located in the north-eastern part of the Tokyo metropolitan area. Its administrative division was formed when the municipality was established in 1953. A *cho* is a fundamental administrative areal unit, and the variation in its geometric attributes is very large. For example, the area of the biggest *cho* is 185 times larger than that of the smallest, and the longest distance from the center of one *cho* to the center of a neighboring *cho* is 22 times longer than the shortest.

The standardized binary-generated distance decay and the first-order neighbors' spatial weight matrices $\boldsymbol{W}_b^*$, $\boldsymbol{W}_d^*$, $\boldsymbol{W}_g^*$, and $\boldsymbol{W}_k^*$ can be derived from the topological attributes of the polygons of the *cho*s and the TIN data, which are built by GIS, e.g., ArcGIS, and a Digital Map 2500 constructed by the Geospatial Authorities of Japan (Can 1996; Zhang 1999; Zhang and Murayama 2000). Then the spatial prominences p_i can be calculated according to (5.7). Figure 5.3 shows four distribution maps of the classified prominences of *cho*s in Matsudo city.

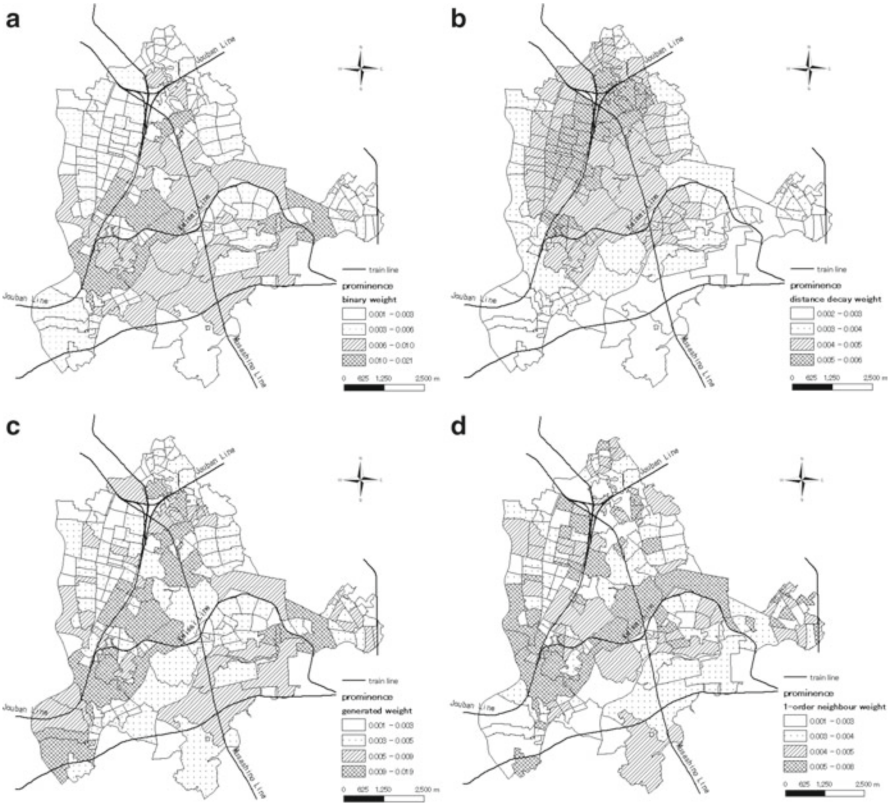


Fig. 5.3 Prominence distributions

Figure 5.3a shows that the prominent *cho*s with a p_i value larger than 0.006 are concentrated on the triangular region formed by the eastern side along the JR Jouban train line, the western side along the JR Musashino train line, and the northern and southern sides along the Keisei subway line. There are many large and non-compact *cho*s in the region. In contrast, small and compact *cho*s show lower prominence values and are located on the periphery of Matsudo city.

Figure 5.3b shows a distribution map of the prominences based on distance decay weight. As shown in this figure, *cho*s with a p_i larger than 0.004 tend to be concentrated in the central and northern regions of the city. There are many small *cho*s with relatively regular shapes. Conversely, many larger *cho*s with a p_i of less than 0.004 are distributed in the southern and western parts of the periphery of the city. The distribution of prominences appears to be more clustered than the other three distributions.

The distribution of prominences based on the generated weight of the *cho*s is represented in Fig. 5.3c, and it can be seen that *cho*s with larger prominence values are distributed along the main train lines where many large and non-compact *cho*s are located. Conversely, in the region with smaller prominence values, many small *cho*s with relatively regular shapes are located. Another large region with a smaller prominence value is located at the southern borders along the Keisei subway line.

As can be seen from the distribution of prominences based on first-order neighbors weight (Fig. 5.3d), no clustered regions are formed, and *cho*s with different prominence values are widely distributed all over Matsudo city.

5.5 Conclusion and Future Prospects

In geography and regional science, an areal unit which has special geometric attributes and maintains a significant spatial correlation and spatial interaction close to adjacent units is called a prominent zone. The prominence of irregular units can be measured by a prominence index, which is a stationary distribution of Markov chain transition matrices which is identical to a spatial weight matrix.

In this chapter, in order to identify the strength of the potential interaction and correlation between spatial units, we reviewed previous research in order to create a spatial weights matrix, and discussed some explicit guidelines on the specifications of the spatial weights matrix and many different types of weighting schemes. Then four definitions of spatial weights were shown, i.e., binary, distance decay, generalized, and *k*-order neighbors weights, and prominences were obtained from these weights matrices. Finally, we used these to analyze the spatial pattern in Matsudo City. The result of this analysis is shown as though different prominences can be obtained from different weights matrices, but a generalized weight matrix is more appropriate for measuring the prominence of units than the distance decay or *k*-order matrix.

As we know, the spatial structure of a city is determined not only by geometric attributes and topological attributes, but also by the social and economic thematic attributes of areal units in city. Therefore it is necessary to create some definitions and measures for prominence units, and use them to analyze urban spatial structures. The definitions proposed began with confirming the relationship between the prominence index and the geometric attributes of areas, and then expanding these to include all the spatial attributes of the areas, i.e., geometric attributes, topological attributes, and thematic attributes. In the future, an approach to implement the definitions will be tested and evaluated to analyze the spatial structure of a city.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Bavaud F (1998) Models for spatial weights: a systematic look. Geogr Anal 30:153–171

Boots BN (1982) Comments on the use of eigenfunctions to measure structural properties of geographic networks. Environ Plann A 14:1063–1072

Can A (1996) Weight matrices and spatial autocorrelation statistics using a topological vector data model. Int J Geogr Inform Syst 10:1009–1017

Cliff AD, Ord JK (1969) The problem of spatial autocorrelation. In: Scott AJ (ed) London papers in regional science, vol 1, Studies in regional science. Pion, London, pp 25–55

Cliff AD, Ord JK (1981) Spatial processes: models and applications. Pion, London

Getis A (2009) Spatial weights matrices. Geogr Anal 41:404–410

Getis A, Aldstadt J (2004) Constructing the spatial weights matrix using a local statistic. Geogr Anal 36:90–104

Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. Geogr Anal 24:189–206

Griffith DA (1996) Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In: Arlinghaus SL et al (eds) Practical handbook of spatial statistics. CRC, Boca Raton, pp 65–82

Griffith DA, Jones KG (1980) Explorations of the relationship between spatial structure and spatial interaction. Environ Plann A 12:187–201

Ord J, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27:286–306

Tinkler KJ (1972) The physical interpretation of eigenfunctions of dichotomous matrices. Trans Inst Br Geogr 55:17–46

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240, Supplement

Zhang C (1999) Development of a spatial analysis tool for irregular zones using the spatial data framework. Geogr Rev Jpn 72:166–177 (in Japanese with English abstract)

Zhang C, Murayama Y (2000) Testing local spatial autocorrelation using *k*-order neighbors. Int J Geogr Inform Sci 14:681–692

Zhang C, Murayama Y (2003) Evaluation on the prominences of irregular areas based on spatial weight matrices. Geogr Rev Jpn 76A:777–787 (in Japanese with English abstract)

Chapter 6
Geographically Weighted Regression in Geospatial Analysis

Rajesh Bahadur Thapa and Ronald C. Estoque

6.1 Introduction

Geographically weighted regression (GWR) is a local spatial statistical technique for exploring spatial non-stationarity. The assumption in GWR is that observations nearby have a greater influence on parameter estimates than observations at a greater distance. This is very close to Tobler's first law of geography—everything is related to everything else, but near things are more related than distant things (Tobler 1970). GWR was developed on the basis of the traditional regression framework which incorporates local spatial relationships into the framework in an intuitive and explicit manner (Brunsdon et al. 1996; Fotheringham and Brunsdon 1999; Fotheringham et al. 2002).

Basically, GWR is based on the non-parametric technique of locally weighted regression developed in statistics for curve-fitting and smoothing applications, where local regression parameters are estimated using subsets of data proximate to a model estimation point in variable space (Wheeler and Paez 2010). The innovation with GWR is using a subset of data proximate to the model calibration location in geographical space instead of variable space. While the emphasis in traditional locally weighted regression in statistics has been on curve-fitting, that is estimating or predicting the response variable, GWR has been presented as a method to

R.B. Thapa (✉)
Earth Observation Research Center, Space Applications Mission Directorate,
Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: rajesh.thapa@jaxa.jp

R.C. Estoque
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Don Mariano Marcos Memorial State University, Bacnotan, La Union, Philippines

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_6, © Springer Japan 2012

conduct inferences on spatially varying relationships in an attempt to extend the original emphasis on prediction to confirmatory analysis (Paez and Wheeler 2009).

Historical evidence suggests that Cleveland and Devlin (1988) introduced locally weighted regression techniques known as locally weighted polynomial regression. At each point in the data set a low-degree polynomial is fitted to a subset of the data, with explanatory variable values near the point whose response is being estimated. The polynomial is fitted using weighted least squares, giving more weight to points near those whose response is being estimated and less weight to points further away. The value of the regression function for the point is then obtained by evaluating the local polynomial using the explanatory variable values for that data point. This is a way of estimating a regression surface through a multivariate smoothing procedure, i.e., fitting a function of the independent variables locally and in a moving fashion which is analogous to how a moving average is computed for a time series. In this framework, with local fitting we can estimate a much wider class of regression surface than with the usual classes of parametric functions, such as polynomials.

In the mid-1990s, the existence of spatial non-stationarity and the essence of its systematic analysis were recognized where a conventional regression model failed to explain the relationships between some sets of variables and observations in space. The nature of the model must alter over space in order to reflect the structure within the data. To address this problem, Brunsdon et al. (1996) developed GWR, which attempts to capture this variation by calibrating a multiple regression model allowing different relationships at different points in space.

6.2 GWR: How Does It Work?

Let's consider a conventional regression model

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \tag{6.1}$$

where y_i is the estimated value of the dependent variable for observation i, β_0 is the intercept, β_k is the parameter estimate for variable k, x_{ik} is the value for the k-th variable for observation i, and ε_i is the error term. In traditional regression, the parameter estimates β_k are assumed to be spatially stationary, but in reality there will be intrinsic differences in relationships over space which may be of a non-stationary character. The non-stationary problem can be measured using GWR (Fotheringham et al. 2002; Platt 2004). Conceptually, the GWR permits the parameter estimates of a multiple linear regression model to vary locally (6.2). GWR extends the conventional regression framework by allowing local rather than global parameters to be estimated. Instead of calibrating a single regression equation, it generates a separate regression equation for each observation. Each equation is calibrated using a different weighting of the observations contained in the data set. The GWR model is written as

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{6.2}$$

where $\beta_0(u_i, v_i)$ denotes the coordinates of the i-th point in space, and $\beta_k(u_i, v_i)$ is a realization of the continuous function at point i (Fotheringham et al. 1998). This allows for a continuous surface of parameter values, and measurements of this surface are taken at certain points to denote the spatial variability of the surface. In (6.2), the parameters are assumed to be spatially invariant. Thus the GWR recognizes that spatial variations in relationships might exist, and provides a way in which they can be measured. The obvious difference between this model (6.2) and the traditional regression model (6.1) is that the regression coefficients are estimated at each data location, whereas in the traditional model they are fixed for the study area.

In GWR, an observation is weighted in accordance with its proximity to location i, so that the weighting of an observation is no longer constant in the calibration, but varies with i. Data from observations close to i are weighted more than data from observations farther away. That is,

$$\hat{\beta}\left(u_i, v_i\right) = (X^T W\left(u_i, v_i\right)X)^{-1} X^T W\left(u_i, v_i\right)y \tag{6.3}$$

where $\hat{\beta}$ represents an estimate of β, $X = (X_1^T, X_2^T, \ldots, X_n^T)^T$ is the design matrix of explanatory variables, which includes a leading column of ones for the intercept, $W_{(u_i, v_i)}$ is the n-by-n diagonal weights matrix calculated for each calibration location, and y is the n-by-n vector of dependent variables. The estimator in (6.3) is a weighted least-squares estimator, but rather than having a constant weight matrix, the weights in GWR vary according to the location of point i. Hence, the weighting matrix has to be computed for each point i, and the weights depict the proximity of each data point to the location of i, with points in closer proximity carrying more weight in the estimation of the parameters for location i.

In GWR, the weight assigned to each observation is based on a distance decay function centered on observation i. The decay function is modified by a bandwidth setting at which distance the weight rapidly approaches zero (Mennis 2006). The bandwidth may be manually chosen by the analyst or optimized using an algorithm that seeks to minimize a cross-validation score given as

$$CV = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq i}(b))^2 \tag{6.4}$$

where n is the number of observations, and $\hat{y}_{\neq i}(b)$ is the fitted value of y_i with the observations for point i omitted from the calculation process so that in areas of sparse observations the model is not calibrated. Plotting the CV score against the required parameter of whatever weighting function is selected will therefore provide guidance on selecting an appropriate value for that parameter. Alternatively, the bandwidth may be chosen by minimizing the Akaike information criteria (AIC) score, which is given as

$$\mathrm{AIC} = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n+\mathrm{tr}(\mathrm{S})}{n-2-\mathrm{tr}(\mathrm{S})}\right\} \tag{6.5}$$

where tr(S) is the trace of the matrix. The AIC method has the advantage of taking into account the fact that the degrees of freedom may vary among models centered on different observations. The AIC have the advantage of being more general in application than the CV statistics because they can be used in Poisson and logistic GWR as well as in linear models. They can also be used to assess whether GWR provides a better fit than a global model, taking into account the different degrees of freedom in the two models (Fotheringham et al. 2002). In addition, the user may choose a fixed bandwidth that is used for every observation, or a variable bandwidth that expands in areas of sparse observations and shrinks in areas of dense observations. Because the regression equation is calibrated independently for each observation, a separate parameter estimate, t-value, and goodness-of-fit is calculated for each observation. These values can thus be mapped, allowing the analyst to visually interpret the spatial distribution of the nature and strength of the relationships among explanatory and dependent variables. For more information on the theory and practical application of GWR, the reader is referred to Fotheringham et al. (2002).

6.3 Empirical Applications

The main output from GWR is a set of location-specific parameter estimates which can be mapped and analyzed to provide information on spatial non-stationarity in relationships. Furthermore, we can estimate local standard errors, derive local t statistics, calculate local goodness-of-fit measures, perform tests to assess the significance of the spatial variation in the local parameter estimates, and run tests to determine if the local model performs better than the global one (Fotheringham et al. 2002). The GWR technique is quite recent, and has only appeared in the scientific literature in the last 15 years. Brunsdon et al. (1996) demonstrated its method and applications in the first case. They investigated the relationships between the rate of car ownership (as the dependent variable) and the proportion of male unemployment and the proportion of households in social class I (as independent variables) in the county of Tyne and Wear in the UK. With the GWR technique, they found that higher rates of car ownership are associated with more rural areas. The rate of car ownership is higher for wards toward the coast and near the southern edge of the region. The relationship is less negative within the highly urbanized core of the study area and southwest to northeast across the southern part of the region.

GWR is calibrated for various applications, and its limitations and opportunities have been discussed widely in the scientific literature. Fotheringham et al. (2002) published a book on GWR which included software. This book provides various fundamental subtopics from basic mechanics to applications of the GWR, with empirical examples and hands-on practice with GWR software. Since then, in the past decade, a number of publications have demonstrated the analytical utility of GWR for investigating a variety of topical areas covering fields from geophysical to socio-economic studies. The following few paragraphs introduce some successful applications of GWR modeling dealing with different problems in various regions world wide.

The relations between riverbank erosion and physiographical variables are assumed to control erosion, which is commonly modeled using regression. For a given river, imagine that a single regression model might be fitted to data on erosion and its geomorphological controls obtained along the length of the river. However, it is likely that the influence of some variables may vary with different geographical locations. In such a case, the stationary regression model is unable to show varying spatial relationships, and therefore should be replaced with a non-stationary model. Atkinson et al. (2003) extended the GWR to predict the binary presence or absence of erosion via the logistic model. This extended model was applied in Afon Dyji in West Wales, UK. The model parameters and the residual deviance of the model varied greatly with distance upstream. The approach presented allowed the inference of spatially varying management practices as a consequence of spatially varying geomorphological processes.

The relationship between the fragmentation of urban development and the forms of administrative and land cover was assessed in order to inform land-use planning in the Roaring Fork/Colorado River corridor of Colorado, USA (Platt 2004). Both conventional regression and GWR were used. While conventional regression provided a good averaged model of change for the entire study area, GWR demonstrated how the process changed locally over space. The results of the global regression showed that the intercept was close to zero, and therefore the fragmentation of urban development was expected to be close to zero in the absence of other forms of fragmentation. The results of the GWR showed that the relationships between changes in the fragmentation of urban development and other fragmentation variables varied significantly within the study area. The analysis also clarified the drivers of fragmentation.

Traditionally, the relationship between tree diameter and total height in a forest stand is investigated by using linear or non-linear regression models in which the spatial heterogeneity in the relationship is largely ignored. Zhang et al. (2004) explored and modeled the spatial variations in the tree diameter–height relationship in a eucalypt stand using GWR. In their modeling, GWR captured spatial variations by calibrating a multiple regression model fitted at each tree, and weighting all neighboring trees by a function of distance from the subject tree. This produced a set of parameter estimates and model statistics (e.g., model R^2) for each tree in the stand. The results indicated that GWR significantly improved model fitting over ordinary least squares (OLS). The GWR model produces smaller model residuals across diameter classes than the traditional OLS model.

Conventional regression models postulating a geographically invariant spatial process have typically been used for relating local variations in disease incidence rates to global association rules. However, ecological associations vary geographically, because the meaning of covariates may change depending on different geographical contexts.

To deal with this problem, Nakaya et al. (2005) extended GWR applications to ecological analysis by introducing an extended statistical tool called GWPR (geographically weighted Poisson regression). This method is a type of conditional kernel regression which uses a spatial weighting function to estimate spatial variations in Poisson regression parameters. An application of the tool is demonstrated based on a regional working-age mortality dataset for the Tokyo metropolitan area, Japan. The results indicate that there are significant spatial variations in the relationships between working-age mortality and occupational segregation, and between working-age mortality and unemployment, throughout the metropolitan area, where conventional models may yield misleading results.

Previous approaches to mapping the results of GWR have primarily employed an equal-step classification and sequential no-hue color scheme for choropleth mapping of parameter estimates. Mennis (2006) showed a cartographic approach which may hinder the exploration of spatial non-stationarity by inadequately illustrating the spatial distribution of the sign, magnitude, and significance of the influence of each explanatory variable on the dependent variable. Approaches for improving the mapping of the results of GWR are illustrated using a case study analysis of population density and median home value relationships in Philadelphia, Pennsylvania, USA. These approaches employ data classification schemes informed by (non-spatial) data distribution, diverging color schemes, and bivariate choropleth mapping.

Mei et al. (2006) demonstrated a mixed GWR model by considering that some coefficients of the explanatory variables are constant, while others vary spatially. Besides the F-approximation, which has frequently been used in the literature of the GWR technique, a statistical inference framework including a bootstrap procedure for deriving the p-value of the test is suggested. The performance of the test is investigated by simulations using hypothetical data. It is demonstrated that both the F-approximation and the bootstrap procedure work satisfactorily.

Tu and Xia (2008) used GWR to examine the relationships between land use and water quality in eastern Massachusetts, USA. The application of GWR models found that the relationships between land use and water quality were not constant over space, but showed great spatial non-stationarity. GWR models were able to reveal information previously ignored by OLS models on the local causes of water pollution, and thereby improve the model's ability to explain the local situation of water quality.

Several studies have indicated that there is a positive relationship between green vegetation land cover and wealthy socio-economic conditions in urban areas. Ogneva-Himmelberger et al. (2009) explored spatial variations in the relationship between socio-economic conditions and green vegetation land cover across urban, suburban, and rural areas in Massachusetts, USA, using GWR. The results showed that there is a considerable spatial variation in the character and strength of the relationship for each model. An impervious surface is generally a strong predictor of the level of wealth, as measured by the variables included in the analysis, at the scale of census block group. However, the strength of the relationship varies geographically.

Revealing spatially varying relationships between urban growth patterns and underlying determinants is important in order to gain a better understanding of the local dimensions of urban development. In a case study of the city of Nanjing, China, Luo and Wei (2009) employed global (conventional regression) and local logistic GWR regressions to model the probability of urban land expansion against a set of spatial variables. The logistic GWR significantly improved the global logistic regression model in terms of a better model goodness-of-fit and a lower level of spatial autocorrelation of residuals. The logistic GWR model allowed the model parameters to vary across space, which provided deep insights into the spatial variations of the urban growth pattern. It demonstrated that the spatial variability of each factor influencing urban land expansion is significant and presents different patterns. Distinctive local patterns and effects of urban growth were found in Nanjing, and these were shaped by local urban spatial and institutional structures. A probability surface of urban growth, which is generated from raster calculations among the parameter and variable surfaces, provides a clear scenario of urban growth patterns, and can be useful for decision making. Although logistic GWR reveals the spatial variations of the influences of spatial variables on urban land expansion more efficiently, the interpretation of such variations should be done carefully and be related to the contextual information about the study area.

Clement et al. (2009) applied GWR to identify drivers of forest transition in Northern Vietnam. The GWR model highlighted the spatial variation of the relationship between the percentage of land afforested and its proximate causes. Factors identified as having a major impact on afforestation are the presence or proximity of a wood-processing industry, the distance to highways, and the land allocation to households. Whereas the former two factors are positively correlated with afforestation in most areas of the province, an unexpected negative correlation was observed for the latter factor. However, an analysis of these results concluded that during the time period considered, afforestation was largely driven by state organizations on protected state-owned land, and forestry was not a significant component of household economic activities.

Jaimes et al. (2010) explored the factors that have induced the loss of forest areas in the State of Mexico from 1993 to 2000 using GWR. The behavior of variables at a local level was analyzed, and some of these were found to present significant spatial variability. This represents an improvement on the understanding offered by global analysis because, rather than producing an average coefficient for the entire territory, this technique yields an estimated coefficient for each location analyzed; in other words, the type of relationship that exists in each portion of the territory is ascertained, and not simply a general overview. This method reveals aspects of the relationships which do not emerge when using traditional global specifications, such as a sign change in some of the parameter estimates.

Despite a growing appreciation of the variation in urban thermal environments and driving factors, relatively little attention has been paid to issues of spatial non-stationarity and scale-dependence, which are intrinsic properties of the urban ecosystem. Li et al. (2010) used GWR to explore the scale-dependent and spatial non-stationary relationships between urban land surface temperature and environmental determinants in the city of Shenzhen, China. The GWR results were compared with those from the OLS (ordinary least squares) method, and it was found that the

GWR model provides a better fit than the traditional OLS model. It also provides detailed local information about the spatial variations of surface temperature which are affected by geographical and ecological factors.

Significant relationships between land use and water quality have been found in watersheds around the world. The relationships are commonly examined by conventional statistical methods, such as OLS regression and Spearman's rank correlation analysis, which assume that the relationships are constant across space. However, the relationships often vary over space because the characteristics of watersheds and the sources of pollution are not the same in different places. Tu (2011) applied GWR to analyze the spatially varying relationships between land use and water quality indicators across watersheds with different levels of urbanization in eastern Massachusetts, USA. This study found that the relationships between water quality and land use, and the abilities of land use indicators to explain water quality, vary across the urbanization gradient in the watersheds. The percentages of commercial and industrial land have stronger positive relationships with the concentrations of water pollutants in less urbanized areas than in highly urbanized areas. The percentages of agricultural land, residential land, and recreational use show significant positive relationships with the concentrations of water pollutants at some sampling sites within less urbanized areas, whereas they have significant negative relationships at some sampling sites within highly urbanized areas. Thus, the adverse impact of land-use changes on water quality is more substantial in less urbanized suburban areas than in highly urbanized city centers.

Landscape fragmentation is usually caused by many different anthropogenic influences and landscape elements. Scientifically revealing the spatial relationships between landscape fragmentation and related factors is very important for land management and urban planning. Former studies on statistical relationships between landscape fragmentation and related factors were almost all global and single-scaled. In fact, landscape fragmentation and its causal factors are usually location-dependent and scale-dependent. Therefore, Gao and Li (2011) used GWR to examine spatially varying and scale-dependent relationships between effective mesh size, an indicator of landscape fragmentation, and related factors in the city of Shenzhen, China. The results show that relationships are spatially non-stationary and scale-dependent, as indicated by the clear spatial patterns of parameter estimates obtained from GWR models. Moreover, GWR models show a better performance than OLS models with the same independent variable.

Dengue fever is one of the most rapidly spreading mosquito-borne viral diseases in the world. In the last 50 years, the incidence of dengue has increased 30-fold, and has extended to new areas across both rural and urban environments (World Health Organization 2009). Unfortunately, the relationships between the incidence of dengue and mosquito abundance, and between the incidence of dengue and human density, are still not well understood (Lin and Wen 2011). The only way for dengue to spread in the human population is through the human–mosquito–human cycle. However, most research in this field discusses dengue–mosquito or dengue–human relationships over a particular study area, and only a few researchers have explored the local spatial variations of dengue–mosquito and dengue–human relationships within a study area. Lin and Wen (2011) used GWR to analyze spatial relationships and identify geographical heterogeneities by using information about entomology and dengue cases in the cities of Kaohsiung and Fengshan, Taiwan. Their findings indicated that dengue–mosquito and dengue–human relationships were significantly spatially non-stationary. This means that in some areas a higher incidence of dengue was associated with higher vector/host densities, but in some areas a higher incidence was related to lower vector/host densities. The GWR model differentiated between the relationships of dengue incidence with immature mosquito and human densities. This study provides insights into the spatial targeting of intervention against dengue epidemics.

6.4 Future Outlook and Concluding Remarks

During the past few decades, there have been significant advances in the field of quantitative geography. The combined effects of the increasing availability of spatial data, advances in analytical methods, and the development of geographical information science has made these possible. In fact, GIS has been instrumental in providing catalysts for the development of appropriate quantitative methods for spatial analysis, in addition to its traditional functions of storing and displaying spatial datasets. Certainly, the advances in GIS have made us more aware and appreciative of the importance of quantitative spatial analysis. Likewise, continuous developments in computer technology and computation have led quantitative spatial analysis to a more advanced level. All of these aspects, in full consideration of the distance–decay effect popularized by Tobler with the first law of geography, have undoubtedly been significant in the evolution of more advanced techniques for quantitative spatial analysis which are similar to GWR (Fotheringham et al. 2002).

Recognizing that global averages of spatial data are not always helpful, and the fact that spatial non-stationarity occurs when a relationship or pattern that applies in one region does not apply in another (Brunsdon 2011), a local statistical technique to analyze spatial variations in relationships, such as GWR, is necessary. GWR is among recent new developments in local spatial analytical techniques, and has been part of the growing trend in GIS towards local analysis. As local statistics are spatial disaggregations of global ones, local analysis intends to understand the spatial data in more detail (Yu and Wei 2004). Since it was first proposed in the late 1990s, GWR has made valuable contributions in quantitative spatial analysis across many areas of geography. As shown in the previous section, it has been implemented in some of the following fields: socio-economic-related studies, e.g., analyzing the relationships between the rate of car ownership (as the dependent variable) and the proportion of male unemployment and the proportion of households; erosion and water-related studies, e.g., the relationships between riverbank erosion and physiographical variables, and the relationships between land use and water quality; urban-related studies, e.g., the relationships between the fragmentation of urban development and forms of administrative and land cover, the relationships between urban growth patterns and underlying determinants, and the relationships between urban land surface temperature and environmental determinants; forestry-related studies, e.g., drivers of forest transition; environmental and health-related problems, e.g., analyzing the relationships between working-age mortality and occupational segregation, and between working-age mortality and unemployment, and exploring the varying spatial relationships of immature mosquitoes and human densities in the incidence of dengue fever.

The idea of applying geographical weights has been extended to a variety of other techniques, in addition to the linear regression framework. This is to produce geographically weighted descriptive statistics, discriminant analysis, logistic regression, Poisson regression, and geographically weighted probit models (Paez and Wheeler 2009). However, there are still issues and debates confronting GWR. Paez and Wheeler (2009) presented and discussed some of these concerns in their review. For example, some of the important issues include (1) the topic of inference regarding the local regression coefficients, (2) the high level of spatial variability encountered by practitioners, a situation that raises concerns that the results may be misleading if not interpreted with caution, and (3) the computational run-time implementations of the techniques. Fortunately, several approaches and diagnostic tools have been developed to overcome the first two issues (see Paez and Wheeler 2009). However, results must continue to be interpreted with caution. Furthermore, the fact that there have been multiple implementations of GWR may mean that the computational run-time may not be a big deal anymore. Currently, aside from the software available from Fotheringham and his co-investigators (http://ncg.nuim.ie/ncg/GWR/), GWR is now incorporated in ArcGIS software. Furthermore, a free and open source implementation of GWR, written by Bivand and Yu (2011), is available as an R package. The basic elements of the GWR technique can also be programmed easily in other languages and software such as Matlab, or in the spatial econometric toolbox developed by LeSage (2010).

However, while GWR has been proven useful in detecting non-stationarity coefficient patterns, the degree of accuracy may not be extremely high. Therefore, whenever possible any conclusions should be supported by other forms of analysis, such as the expansion method and/or multilevel or hierarchical Bayesian models (Paez and Wheeler 2009). While this requires extra effort and time, it would help to increase the level of confidence with respect to any conclusions drawn from the analysis. Nevertheless, its strength in capturing the spatial relationships of given variables at the local level, based on the concept of non-stationarity, is indeed its most important contribution to the quantitative revolution of spatial analysis. Its applications will certainly continue to expand across the diverse fields of quantitative geography.

References

Atkinson PM, German SE, Sear DA, Clark MJ (2003) Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. Geogr Anal 35:58–82

Bivand R, Yu D (2011) Package 'spgwr'. http://cran.r-project.org/web/packages/spgwr/spgwr.pdf. Accessed 6 July 2011

Brunsdon C (2011) Geographically weighted analysis: review and prospect. http://xweb.geos.ed.ac.uk/~gisteac/eeo_agi/2010…/7_brunsdon_11022011.pdf. Accessed 22 July 2011

Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial non-stationarity. Geogr Anal 28:281–298

Clement F, Orange D, Williams M, Mulley C, Epprecht M (2009) Drivers of afforestation in Northern Vietnam: assessing local variations using geographically weighted regression. Appl Geogr 29:561–576

Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83:596–610

Fotheringham AS, Brunsdon C (1999) Local forms of spatial analysis. Geogr Anal 31:340–358

Fotheringham AS, Brunsdon C, Charlton ME (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plann A 30:1905–1927

Fotheringham AS, Brunsdon C, Charlton ME (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester

Gao J, Li S (2011) Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Appl Geogr 31:292–302

Jaimes NBP, Sendra JB, Delgado MJ, Plata RF (2010) Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. Appl Geogr 30:576–591

LeSage J (2010) Econometrics toolbox for MATLAB. http://www.spatial-econometrics.com

Li S, Zhao Z, Miaomiao X, Wang Y (2010) Investigating spatial non-stationary and scale-dependent relationships between urban surface temperature and environmental factors using geographically weighted regression. Environ Model Software 25:1789–1800

Lin CH, Wen TH (2011) Using Geographically Weighted Regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of Dengue. Int J Environ Res Public Health 8:2798–2815

Luo J, Wei YHD (2009) Modeling spatial variations of urban growth patterns in Chinese cities: the case of Nanjing. Landsc Urban Plann 91:51–64

Mei CL, Wang N, Zhang WX (2006) Testing the importance of the explanatory variables in a mixed geographically weighted regression model. Environ Plann A 38:587–598

Mennis J (2006) Mapping the results of geographically weighted regression. The Cartographic Journal 43:171–179

Nakaya T, Fotheringham AS, Brunsdon C, Charlton M (2005) Geographically weighted Poisson regression for disease association mapping. Stat Med 24:2695–2717

Ogneva-Himmelberger Y, Pearsall H, Rakshit R (2009) Concrete evidence & geographically weighted regression: a regional analysis of wealth and the land cover in Massachusetts. Appl Geogr 29:478–487

Paez A, Wheeler DC (2009) Geographically weighted regression. In: Kitchin R, Thrift N (eds.) International Encyclopedia of Human Geography 1:407–414. Oxford: Elsevier

Platt RV (2004) Global and local analysis of fragmentation in a mountain region of Colorado. Agric Ecosyst Environ 101:207–218

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Geogr Anal 46:234–240

Tu J (2011) Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. Appl Geogr 31:376–392

Tu J, Xia ZG (2008) Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. Sci Total Environ 407:358–378

Wheeler DC, Paez A (2010) Geographically weighted regression. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis: software tools, methods and applications. Springer, Berlin, pp 461–486

World Health Organization (WHO) (2009) Dengue: guidelines for diagnosis, treatment, prevention and control, New Edition. WHO, Geneva

Yu D, Wei YD (2004) Geographically weighted regression: investigation of spatially varying relationships: methods, techniques, and implementation. University of Wisconsin, Milwaukee, Paper presented during the GIS Day

Zhang L, Bib H, Cheng P, Davis CJ (2004) Modeling spatial variation in tree diameter–height relationships. For Ecol Manage 189:317–329

Chapter 7
Weight of Evidence in Geospatial Analysis

Rajesh Bahadur Thapa

7.1 Introduction

Weight of evidence (WofE) is a quantitative method for combining evidence in support of a hypothesis. An evidence-based approach involves an assessment of the relative values of different pieces of information that have been collected in previous steps. ECHA (2010) defines WofE as "the process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion." A representative value needs to be assigned to each piece of information using a formalized weighting procedure. The evidence can be called as a factor, and can often influence the weight given owing to the quality of the data, the consistency of results, the nature and severity of effects, and the relevance of the information.

WofE is entirely based on the Bayesian approach of conditional probability. This method combines spatial data from a variety of sources to describe and analyze interactions, provides evidence for decision making, and makes predictive models. Basically, the method concerns the probability of detecting a certain event, which could be a given category of land-use change, for example, possibly an event of land-use change from agricultural area to built-up surface in relation to potential evidence (proximity to urban centers, roads, water, etc.), often called the driving factors of change (Thapa and Murayama 2011).

Historical evidence shows that Peirce (1878) was very close to the best definition of WofE, namely the logarithm of a Bayes factor, which is the ratio of the posterior to the prior odds. A theoretical expression and discussion of WofE modeling at an earlier stage can be found in Good (1950, 1979). This method was originally developed

R.B. Thapa (✉)
Earth Observation Research Center, Space Applications Mission Directorate,
Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan

Formerly in Division of Spatial Information Science, Graduate School of Life
and Environmental Sciences, University of Tsukuba,Tsukuba, Japan
e-mail: rajesh.thapa@jaxa.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_7,

for a non-spatial application, and therefore its applications dominated the literature in statistics and medical related fields until the mid-1980s. Its application to medical fields was promising, for instance, when the evidence consisted of a set of symptoms, and the hypothesis was "this patient has disease x." For each symptom, a pair of weights was calculated, one for the presence of the symptom and one for the absence of the symptom. The magnitude of the weights depended on a measured association between the symptom and the occurrence of the disease in a large group of patients. The weights could then be used to estimate the probability that a new patient would get the disease, based on the presence or absence of symptoms (Spiegelhalter 1986; Raines et al. 2000). In this chapter, I briefly review the progress in the methodology and applications of WofE in the field of geospatial analysis, discuss how the method works, and conclude with an outlook for the future.

7.2 Methodology and Applications in Geospatial Analysis

Since the late 1980s, the geoscience field has adopted WofE modeling for geospatial applications, the earliest of which were maps of mineral potential (Bonham-Carter et al. 1988). A pattern of mineral deposits is related to several map layers representing geological data that may be indicative of the occurrence of mineral deposits (Bonham-Carter 1994). The method has gradually been integrated with spatial databases, and has been used for a variety of purposes in various geographic regions. Aspinall (1992) described an inductive modeling procedure integrated with geographical information science (GIS) and the Bayesian theorem for wildlife habitat mapping. The use of the modeling procedure is illustrated through an analysis of the winter habitat relationships of red deer in the Grampian Region, north–east Scotland. The habitat data sets used to construct the model were the accumulated frost and altitude records obtained from maps, and land cover derived from satellite imagery.

Bonham-Carter (1994) illustrated the modeling process in a probabilistic framework, so that the weighting of individual map layers was based on a Bayesian probability model. In particular, the WofE model was presented in a map context, with examples showing applications to mineral-potential mapping in Meguma terrane, Nova Scotia, Canada. The relationships of WofE to the methods used in the expert system of the prospectors are explained with a very simple example of the system's inference network.

Similarly, Cheng and Agterberg (1999) proposed a new approach to the WofE method based on fuzzy sets and fuzzy probabilities for mineral-potential mapping. This approach can be considered as a generalization of the ordinary weights of evidence method, which is based on binary patterns of evidence and has been used in conjunction with GIS for mineral-potential mapping. In the newly proposed method, instead of separating the evidence into binary form, fuzzy sets containing more subjective genetic elements are created; fuzzy probabilities are defined to construct a model for calculating the posterior probability of a unit area containing mineral deposits on the basis of the fuzzy evidence for the unit area. This method can be treated as a hybrid method, which allows objective or subjective definitions of a fuzzy membership function of evidence augmented by an objective definition of fuzzy or conditional probabilities. The posterior probabilities calculated by this method would depend on existing data in a totally data-driven approach, but would also depend partly on expert knowledge when the hybrid method is used.

An ArcView GIS extension of WofE, i.e., Arc-WofE, is publicly available and has applications to mineral potential (Kemp et al. 1999). This enhances further potential uses of WofE in the geospatial field. The system has four core steps: building a spatial digital database; extracting predictive evidence for a particular deposit type based on an exploration model; calculating weights for each predictive map or evidential theme; combining the evidential themes to make a prediction. This extension also provides an expert approach to weighting which can be used when no training points are available (Raines et al. 2000).

The popularity of the WofE method with more geospatial applications was further expanded in other environmental study fields in the first decade of the twenty-first century. A structure for simulating land-use change using the elementary probabilistic methods of the WofE approach was proposed by Almeida et al. (2003). The model framework has been applied to Bauru town in Brazil. This showed how various socio-economic and infrastructural factors can be combined using the WofE approach, which then enables them to predict the probability of changes between land-use types in different cells of the system. Another study over a larger area, which was also conducted in Brazil by Soares-Filho et al. (2004), shows an application of the WofE method to select the most important variables needed for land-cover change analysis and to quantify their influences on each type of land-use transition, e.g., deforestation, land abandonment, and re-growth clearing. They developed a land-cover change simulation model that is responsive to road paving and policy intervention scenarios in central Amazonia. The model assesses the impacts of road paving within the population, as well as policy intervention scenarios.

Romero-Calcerrada and Luque (2006) focused on boreal forest landscapes, and explored a multicriteria approach by using a predictive habitat suitability model for the three-toed woodpecker (*Picoides tridactylus*) based on WofE. Since the method depends on the indicator species which is used as a surrogate of biodiversity value, it can be applied to assess the biodiversity conditions of both managed and protected areas to help decision-making concerning the protection of valuable habitats. Thus, a map of habitat suitability representing a range of probabilities of occurrence offers an objective framework for evaluating the outcomes of different scenarios. Similarly, an objective assessment of habitat suitability provides a rational basis for management decisions incorporating the impact on species habitat. Romero-Calcerrada and Millington (2007) used both WofE and logistic regression to analyze the natural and human factors that contribute to wildfire on the Iberian Peninsula. Unlike expert knowledge approaches to modeling, the WofE approach derives the probabilities of fire occurrence based on the association between mapped occurrences and spatial evidence layers of biophysical data.

An application of the WofE method to perform a vulnerability assessment for the occurrence of elevated nitrate concentrations in the aquifer of Milan, Italy, is given by Masetti et al. (2007). A comparison between the spatial distribution of vulnerability classes and the frequency of occurrences of nitrate in wells shows a high degree of correlation for both low and high nitrate concentrations. Groundwater-specific vulnerability was classified in terms of vulnerability classes and, according to the outcomes of the model, the population density can be considered to be the source of the greatest impact of nitrate. Mean annual irrigation and groundwater depth can be identified as influencing factors in the distribution of nitrate, while agricultural practice appears to be a negligible factor.

Dahal et al. (2008) and Pradhan et al. (2010) presented WofE modeling applications to landslide susceptibility mapping. The former applied the modeling to small catchments of Shikoku, Japan, while the latter applied it to a tropical hilly area in Malaysia. The Japanese case showed the usefulness and capability of the modeling in a small catchment area with a high-resolution data set. The Malaysian case showed the method of calculating the rating factor, and reported that the landslide susceptibility map and the verification results achieved a high predictive accuracy for the model.

Dilts et al. (2009) used WofE techniques to model spatial patterns of wildfire occurrence in relation to landscape-scale drivers of fire in Lincoln County, Nevada, USA. The spatial data sets which were used as potential predictors of fire occurrence included biophysical and socio-economic data. Models were developed and tested for lightning-caused fires over the entire county, and also in forested areas only. Higher fire density and higher lightning-strike density were observed in the eastern half of the county compared with the western half. Overall, the spatial distribution of wildfire occurrence was controlled more by ignition mechanisms than by processes influencing fuel moisture, accumulation, or both.

A recent application of WofE is found in urban growth modeling. Thapa and Murayama (2011) adopted the WofE method integrated in a cellular automata framework for predicting the future spatial patterns of urban growth in the Kathmandu metropolitan region. The model was validated by achieving a highly accurate prediction of urban development patterns for the future under the current scenario across the metropolitan region. Depending on local characteristics and land-use transition rates, the model produced a noticeable spatial pattern of changes in the region. The application of WofE to urban growth modeling can be found in Chap. 13.

7.3 WofE Model: How Does It Work?

Let's now consider a landscape (Fig. 7.1) which has three spatial patterns: forest, road buffer, and an area which is changing from forest to non-forest land. Landscape change is observed along the road network, so the road is considered to be the major driver of the change. If we considered this in binary terms, area change is represented as 1 and no change as 0. Similar assumptions can be made for the road layer, i.e., inside the road buffer as 1 and outside as 0. In this particular case, the WofE concerns the probability of detecting land change (deforestation) influenced by the driver (road).

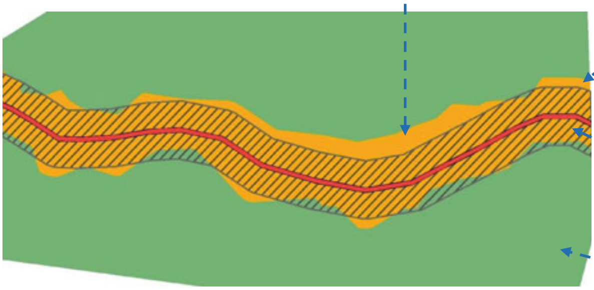


Fig. 7.1 Schematic drawing to illustrate the WofE method

To understand this deforestation process and detect the probability using the WofE technique, the WofE model is synthesized from Bonham-Carter (1994). The areas of deforestation, *D* (landscape change from forest to non-forest), and the explanatory variable, *E* (road buffer), are known, and then the probability of locating the occurrence of deforestation given the presence of the explanatory variable can be expressed by the conditional probabilities given in (7.1).

$$P(D \mid E) = \frac{P(D \cap E)}{P(E)} \tag{7.1}$$

The symbol $\cap$ is a logical intersection or Boolean AND operation. The conditional probability of *D* occurring given the presence of *E* is written as $P(D \mid E)$. Thus, the probability of a deforestation pattern (*D*) occurring given the presence of explanatory variable (*E*) can be expressed as a probability ratio, which follows from the basic definition of conditional probability, followed by the substitution of area proportions as estimates of probabilities, and finally as a ratio of areas. The probability of land change in the WofE modeling framework is expressed as *odds*. *Odds* (*O*) are defined as $O = P/(1-P)$, a ratio of the probability that an event will or will not occur (Bonham-Carter 1994). Now, (7.1) can be converted into odds (7.2).

$$O(D \mid E) = \frac{P(D \mid E)}{1 - P(D \mid E)} = \frac{P(D \mid E)}{P(\bar{D} \mid E)} \tag{7.2}$$

where $\bar{D}$ represents the absence of deforestation, i.e., no change occurred in the landscape. $1 - P(D \mid E)$ becomes $P(\bar{D} \mid E)$ when we consider the probability of *D* (deforestation) being absent given the presence of *E* (explanatory variable, road). *Odds* values of less than 1 correspond to probabilities less than 0.5, and very small probabilities are nearly the same as *odds*. Accordingly, a similar argument is used to derive an expression for the conditional odds of *D* given the absence of *E*:

$$O(D \mid \bar{E}) = \frac{P(D \mid \bar{E})}{P(\bar{D} \mid \bar{E})} \tag{7.3}$$

The WofE method can combine several explanatory variables to predict similar patterns of land change. A pair of weights W^+ (presence) and W^- (absence) can be determined for each predictor pattern (road, and other predictor if any), depending on the measured spatial association with the pattern of land change. The weights may be combined from each pattern to make a predictive map for the change. Taking a single predictor pattern, D, the positive weight W^+ and the negative weight W^- can be expressed as the difference between the prior and posterior logit of D, as follows:

$$W^+ = \ln O\left(D \mid E\right) - \ln O(D) = \ln\left[\frac{O\left(D \mid E\right)}{O(D)}\right] \tag{7.4}$$

$$W^- = \ln O\left(D \mid \bar{E}\right) - \ln O(D) = \ln\left[\frac{O\left(D \mid \bar{E}\right)}{O(D)}\right] \tag{7.5}$$

The WofE method uses the natural logarithm of *odds*, known as *log odds* or *logit*. The logit scale is centered at 0, corresponding to a probability of 0.5, with negative values for *odds* less than 1 and positive values for *odds* greater than 1. After computing the weights, the posterior logit can be generated using the following equations:

$$\ln O\left(D \mid E\right) = \ln O(D) + W^+ \tag{7.6}$$

$$\ln O\left(D \mid \bar{E}\right) = \ln O(D) + W^- \tag{7.7}$$

More explanatory variables can be incorporated with an assumption that the variables are conditionally independent with respect to land change (i.e., deforestation). The following expression can be written for more explanatory variables:

$$\ln O(D \mid E_1^k \cap E_2^k \cap \ldots E_n^k) = \ln O(D) + \sum_{j=i}^{n} W_{E_j}^k \tag{7.8}$$

where the superscript k is positive (+) or negative (−) depending on whether the explanatory variable is present or absent, respectively.

The explanatory variables are usually either discrete (e.g., a land-use planning map or other socio-economic data) or continuous (e.g., a slope, proximity to road, river, etc.). Continuous variables need to be transformed into discrete variables. Thence, each explanatory theme has $k = 1,2,\ldots,m$ discrete class values or states which can be associated with weights in respect to the occurrence of events.

The WofE application on landslide susceptibility mapping synthesized from Dahal et al. (2008) is presented as an example. The application test area, with less than 400 ha of the Moriyuki catchment, is located in the northeast of the Shikoku region, Japan. The catchment had as many as 76 landslides (Fig. 7.2) due to heavy rainfall which occurred in October 2004. Thematic maps, i.e., slope, aspect, relief,

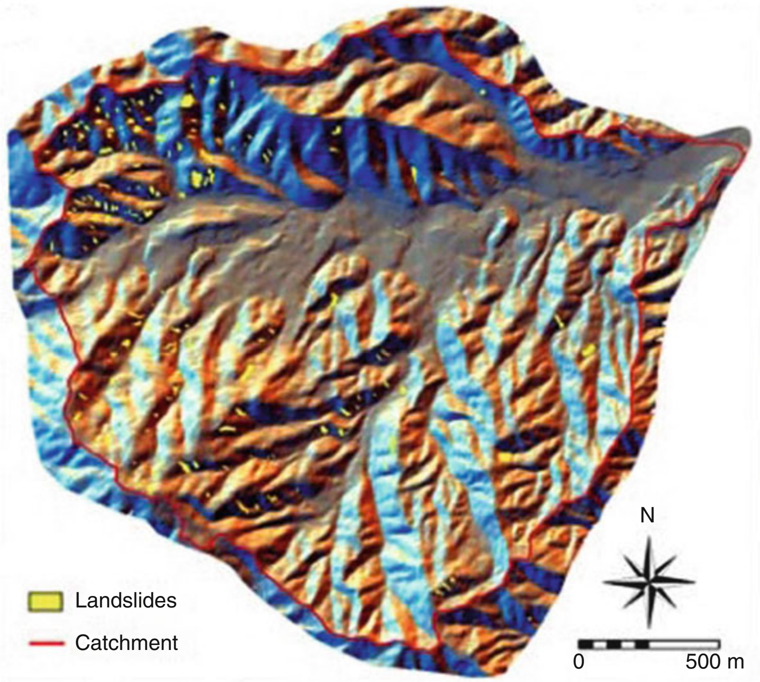


Fig. 7.2 Landslide events in the Moriyuki catchment (reproduced from Dahal et al. 2008)

flow accumulation, soil depth, soil type, land use, and distance to road, were used as landslide predictive factors. The size of the landslide and the number of potential events varies in different landscapes depending on local geo-environmental characteristics. Therefore, each thematic map was logically classified into various category ranges in order to estimate geographically varying weights. The maps of categorical factors were overlaid with the landslide event map, and this produced weights for each map using WofE techniques. The weighted factor maps were linearly combined to create a landslide susceptibility index map. The index map was cross-validated with a landslide event map and showed a considerable success rate, i.e., 80.7% of the WofE-based predictive model. Based on the success rate ratio, a map with five landslide susceptibility zones, i.e., very low, low, moderate, high, and very high, was established and is shown in Fig. 7.3.

7.4 Future Outlook

From the discussion and examples above, it is clear that the WofE method can combine spatial data from diverse sources to describe and analyze interactions, provide support for decision makers, and make predictive models. The statistical association between an event and the associated factors determines the weights. The WofE

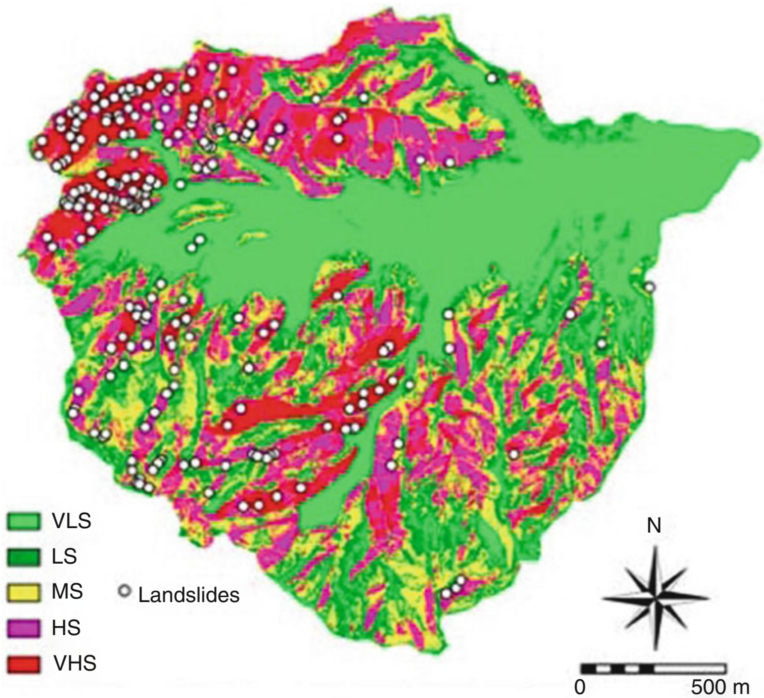


Fig. 7.3 Landslide susceptibility map of the Moriyuki catchment. *VHS* very high susceptibility, *HS* high susceptibility, *MS* moderate susceptibility, *LS* low susceptibility, *VLS* very low susceptibility (reproduced from Dahal et al. 2008)

method itself is combined with Bayes' rule of probability, with an assumption of conditional independence. The model is given in log–linear form so that the weights from the evidential themes can be added.

In this method, weight values are easy to interpret. A positive weight for a particular evidential-theme indicates that a larger proportion matched the conditions of that theme than would occur due to chance, whereas the inverse is true for negative weights. A weight of zero indicates that the training points are spatially uncorrelated to the theme. The range in weight values for a particular evidential theme gives an overall measure of how important the theme is for modeling. Uncertainties due to variances of weights and missing data allow the relative uncertainty in posterior probability to be estimated and mapped. Because conditional independence is never completely satisfied, the posterior probabilities are usually overestimated in absolute terms. However, the relative variations in posterior probability (as observed in spatial patterns on the response map) are usually not much affected by violations of this assumption, as stated by Raines et al. (2000).

Being a Bayesian method, it has a number of advantages over other more traditional spatial modeling methods. Much appreciated are the ability to accommodate both categorical and ordered data, the robustness to small sample sizes, and the ability to use data that are not normally distributed (Bonham-Carter et al. 1988). As we have observed several applications in the geospatial field, i.e., mineral-potential mapping, landslide susceptibility, habitat suitability, wildfire, land-use and land-cover change analysis, and urban growth modeling, the future of this method is emerging. It may be possible that the wider applications of WofE will soon be used to solve global environmental problems, such as in REDD+.

References

Almeida CM, Batty M, Monteiro AVM, Camara G, Soares-Filho BS, Ceroueira GC, Pennachin CL (2003) Stochastic cellular automata modelling of urban land use dynamics: empirical development and estimation. Comput Environ Urban Syst 27:481–509

Aspinall RJ (1992) An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data. Int J Geogr Inform Syst 6:105–121

Bonham-Carter G (1994) Geographic information systems for geoscientists: modeling with GIS. Pergamon, New York

Bonham-Carter GF, Agterberg FP, Wright DF (1988) Integration of geological datasets for gold exploration in Nova Scotia. Am Soc Photogram Rem Sens 54:1585–1592

Cheng Q, Agterberg FP (1999) Fuzzy weights of evidence method and its application in mineral potential mapping. Nat Resour Res 8:27–35

Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. Environ Geol 54:311–324

Dilts TE, Sibold JS, Biondi F (2009) A weights-of-evidence model for mapping the probability of fire occurrence in lincoln county, Nevada. Ann Assoc Am Geogr 99:712–727

ECHA (European chemical Agency) (2010) Practical guide 2: how to report weight of evidence. http://echa.europa.eu/. Accessed 4 Mar 2011

Good IJ (1950) Probability and the weighing of evidence. C. Griffin, London

Good IJ (1979) Studies in the history of probability and statistics: A. M. Turing's statistical work in World War II. Biometrika 66:393–396

Kemp LD, Bonham-Carter GF, Raines GL (1999) Arc-WofE: arcview extension for weights of evidence mapping. http://www.ige.unicamp.br/wofe. Accessed 7 Mar 2011

Masetti M, Poli S, Sterlacchini S (2007) The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination. Nat Resour Res 16:109–119

Peirce CS (1878) The probability of induction. Popular Science Monthly 12:705–718. Reprinted (1956) in Newman JR (ed) The world of mathematics, vol 2. Simon and Schuster, New York, pp 1341–1354

Pradhan B, Oh H, Buchroithner M (2010) Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. Geomat Nat Hazards Risk 1:199–233

Raines GL, Bonham–Carter G, Kemp L (2000) Predictive probabilistic modeling: using ArcView GIS. ArcUser, April–June, pp 45–48. http://www.esri.com

Romero-Calcerrada R, Luque S (2006) Habitat quality assessment using weights-of-evidence based GIS modelling: the case of *Picoides tridactylus* as species indicator of the biodiversity value of the Finnish forest. Ecol Model 196:62–76

Romero-Calcerrada R, Millington JDA (2007) Spatial analysis of patterns and causes of fire ignition probabilities using logistic regression and weights-of-evidence based GIS modeling. Geophys Res Abstr 9:01337

Soares-Filho BS, Alencar A, Nespad D, Cerqueira GC, Dial M, Del C, Solozarno L, Voll E (2004) Simulating the response of land-cover changes to road paving and governance along a major Amazon Highway: the Santarem–Cuiaba corridor. Glob Chang Biol 10:745–764

Spiegelhalter DJ (1986) Probabilistic prediction in patient management and clinical trials. Stat Med 5:421–33

Thapa RB, Murayama Y (2011) Urban growth modeling of Kathmandu metropolitan region, Nepal. Comput Environ Urban Syst 35:25–34

Chapter 8
Markov–Cellular Automata in Geospatial Analysis

Courage Kamusoko

8.1 Introduction

Spatial simulation models are indispensable for modeling land use/cover changes (Wu and Webster 1998; Messina and Walsh 2001; Soares-Filho et al. 2002), deforestation and land degradation (Lambin 1994; Lambin 1997; Etter et al. 2006; Moreno et al. 2007), urban growth (Clarke et al. 1997; Couclelis 1989; Cheng and Masser 2004; Gar-On Yeh and Li 2009), climate change (Dale 1997) and hydrology (Matheussen et al. 2000). For land use/cover change studies, spatial simulation models are critical for understanding the driving forces of change, as well as to produce "what if" scenarios that can be used to gain insights into future land use/cover changes (Pijanowski et al. 2002; Eastman et al. 2005; Torrens 2006). Recently, the knowledge domain of spatial simulation modeling has advanced owing to the rapid developments in computer technology, coupled with the decrease in the cost of computer hardware. In addition, developments in geospatial, natural and social sciences concerning bottom-up, dynamic and flexible self-organizing modeling systems, complemented by theories that emphasize the way in which decisions made locally give rise to global patterns, have enriched spatial simulation models (Tobler 1979; Wolfram 1984; Couclelis 1985; Engelen 1988; Wu and Webster 1998; Batty 1998). To date, numerous spatial simulation models have been developed and applied, particularly for land use/cover modeling (Clarke et al. 1997; Kaimowitz and Angelsen 1998; Messina and Walsh 2001; Soares-Filho et al. 2002; Walsh et al. 2006).

C. Kamusoko (✉)
Overseas Operations Division, Asia Air Survey Co. Ltd.,
Tokyo, Japan

Formerly in Integrative Environmental Sciences Program, University of Tsukuba,
Tsukuba, Japan
e-mail: cou.kamusoko@ajiko.co.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_8,

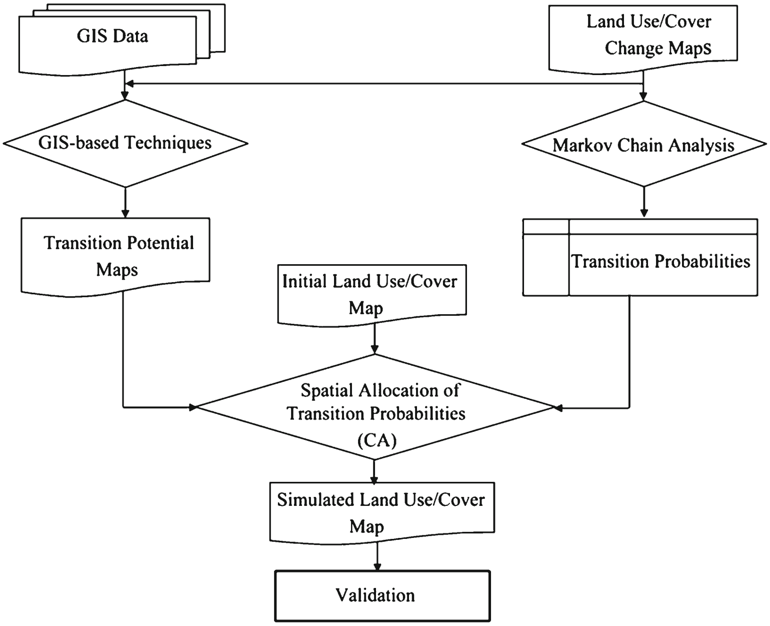


Fig. 8.1 Conceptual framework of the Markov–cellular automata (MCA) model

While a literature review reveals a plethora of spatial simulation models based on different modeling techniques and traditions (Parker et al. 2003; Verburg et al. 2004), in this chapter we focus on the Markov–cellular automata (MCA) model that integrates cellular automata (CA) procedures, Markov chains and geographical information science (GIS)-based techniques such as weight of evidence (WofE) and multi-criteria evaluation (MCE) (Fig. 8.1). The objective of this chapter is to review the methodological developments of the MCA model. The chapter is organised into five sections. Section 8.2 focuses briefly on the conceptual framework of the MCA model, paying special attention to the basics of Markov chains, GIS-based techniques such as WofE and MCE, and CA models. The application of MCA models in previous studies is described in Sect. 8.3, while Sect. 8.4 focuses on the current status and future prospects of the MCA modeling framework. Finally, Sect. 8.5 gives a summary and conclusions.

8.2 Conceptual Framework of the MCA Model

A MCA model is a spatial model for simulating land use/cover changes in landscapes where land use/cover is viewed as a mosaic of discrete states, and changes are multi-directional (e.g. forest to non-forest or vice versa) (Silverton et al. 1992; Li and Reynolds 1997). In order to gain insights into land use/cover changes in a given landscape, modeling approaches that can adequately represent state-and-transition systems should be used. The MCA model that combines CA with Markov chain analysis and GIS-based techniques (Fig. 8.1) can be used for modeling land use/cover changes since it can effectively represent state-and-transition systems. The Markov chain process uses transition probabilities to control temporal dynamics among the land use/cover classes. Spatial dynamics are controlled by local rules determined either by the CA mechanism (neighborhood configuration) or by its association with the transition potential maps computed from WofE and MCE techniques. The MCA model allows the transition probabilities of one pixel to be a function of neighboring pixels because the CA model consists of a regular grid of cells, each of which can be in one of a finite number of possible states which are updated synchronously in discrete time steps according to a local interaction rule (Messina and Walsh 2001). The transition probabilities of the CA model depend on the state of a cell, the state of its surrounding cells, and the weights associated with the neighborhood context of the cell (White and Engelen 1997). The MCA model (Li and Reynolds 1997) can be expressed as

$$\begin{array}{ll} C(i,j) = m \rightarrow k, & \text{if } R > P_{m,k} \bullet N_k / 4 \\ \text{No change,} & \text{if } R \le P_{m,k} \bullet N_k / 4 \end{array} \tag{8.1}$$

where C (i, j) is the use/cover class of cell (i, j), R is a random number with a uniform distribution, $P_{m,k}$ is the transition probability from one land use/cover class m to k, and N_k is the number of neighboring cells of land use/cover k, which includes the evaluation score of land use/cover transition potential at location i, j. The weight of four is used in (8.1) because each cell is assumed to have four neighbors.

Land use/cover change modeling approaches such as MCA generally consist of three major components: (1) a change demand submodel, (2) a transition potential submodel, and (3) a change allocation submodel (Eastman et al. 2005). The change demand submodel estimates the rate of change between two land use/cover maps from different periods. The results are summarised in a transition probability matrix that expresses the rate of conversion from one land use/cover class to another (Table 8.1). The transition potential submodel determines the likelihood (which can also be expressed as suitability or probability) that land would change from one land use/cover class to another based on biophysical and socio-economic factors (Fig. 8.2). Specifically, it establishes the degree to which locations might potentially change in a future period of time (Eastman et al. 2005). Finally, the change allocation submodel is concerned with the decisions by which specific areas will change, given

Table 8.1 Example of land use/cover change transition probabilities (1993–2000)

		2000		
		Current forest	Unstocked forest	Non-forest
1993	Current forest	**0.85**	0.13	0.02
	Unstocked forest	0.01	**0.94**	0.05
	Non-forest	0.05	0.65	**0.30**

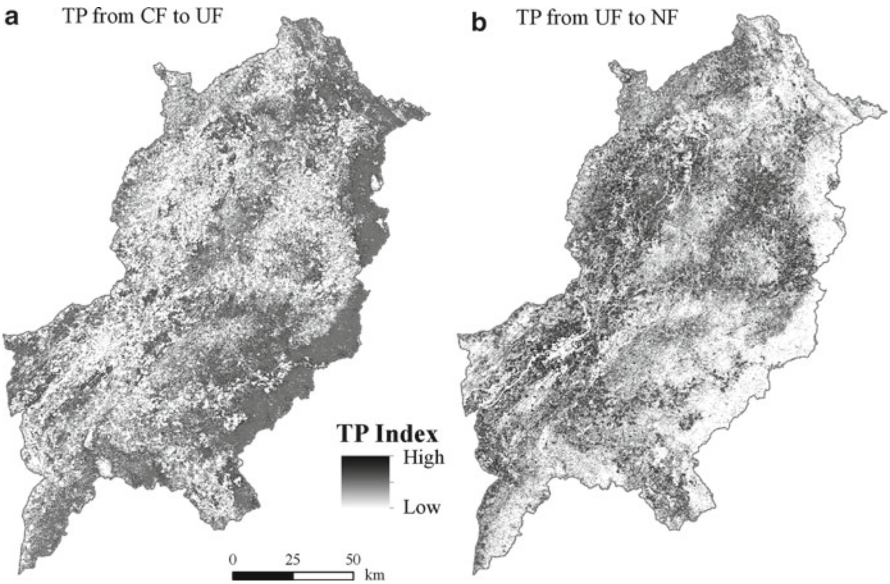


Fig. 8.2 Example of transition potential (TP) maps of Luangprabang province, Lao Peoples' Democratic Republic: (**a**) current forest (CF) to unstocked forest (UF), and (**b**) unstocked forest (UF) to non-forest (NF)

the demand and potential surfaces. In the MCA modeling approach, the change demand model is represented by the Markov chains, while the transition potential and change allocation submodels are represented by GIS-based techniques such as the WofE and MCE models and the CA models, respectively.

In the following subsections, we focus briefly on (1) the computation of transition probabilities using Markov chains, (2) the computation of land use/cover transition potential maps based on WofE and MCE techniques, (3) the spatial allocation of simulated land use/cover probabilities based on a CA model, and (4) the mechanism of the MCA model.

8.2.1 Markov Chain Modeling of Land Use/Cover Changes

Markov chains have been widely used to model land use/cover changes (Drewett 1969; Bell 1975; Bell and Hinojosa 1977; Robinson 1978; Jahan 1986; Muller and Middleton 1994; Wood et al. 2004). A Markov chain is a stochastic model based on transition probabilities which describes a process that moves in a sequence of steps through a set of states (Wu et al. 2006). In essence, the probability that the system will be in a given state at a given time (t_2) is derived from the knowledge of its state at any earlier time (t_1), and does not depend on the history of the system before time t_1 (Petit et al. 2001). This is known as a first-order Markov chain process. The Markov chain can be characterized as stationary or homogeneous in time if the transition probabilities depend only on the time-interval t (i.e. $\Delta t = t_2 - t_1$), and if the time period at which the process is examined is of no relevance (Karlin and Taylor 1975). To model land use/cover change using the stationary and first-order Markov chain, the land use/cover distribution at t_2 is calculated from the initial land use/cover distribution at t_1 based on the transition matrix (Lambin 1994; Petit et al. 2001). The Markov chains can be expressed as

$$v_{t2} = M * v_{t1} \tag{8.2}$$

where v_{t2} is the output land use/cover proportion column vector, v_{t1} is the input land use/cover class proportion column vector, and M is an $m*m$ transition matrix for the time interval $\Delta t = t_2 - t_1$.

In order to model land use/cover changes using Markov chains, it is essential to understand their basic assumptions and limitations. First, land use/cover changes are considered as a stochastic process where the transition probabilities are stationary (homogeneity) and land use/cover classes are in different states of the Markov chain (Wu et al. 2006). However, it is difficult to expect stationarity in transition probabilities because land use/cover changes are the result of the complex dynamics of socio-economic, political and biophysical factors that change over time (Lambin et al. 2000). While the departure from the simple assumptions of stationary, first-order Markov chains is conceptually possible, analytical and computational difficulties emerge. Nonetheless, it might be practical to assume transition probabilities to be stationary if the time span is not too long (Weng 2002). Second, the Markov chains that handle stationary processes are not appropriate for incorporating human activities (Boerner et al. 1996; Weng 2002). Third, the available land use/cover data may be insufficient to estimate reliable transition probabilities (Pastor et al. 1993), particularly in landscapes experiencing rapid land use/cover changes (Wood et al. 2004). Finally, a stochastic Markov chain model does not consider spatial knowledge within each land use/cover class (Boerner et al. 1996).

While the Markov chains have some limitations, they are relatively easy to derive (or infer) from land use/cover data (Wood et al. 2004). Despite the fact that the Markov chains do not reveal the underlying land use/cover change processes, they give the direction and magnitude of change, and that is potentially of use for simulating land use/cover changes (Weng 2002). In addition, the computational requirements of Markov chain models are quite modest.

Table 8.1 shows the forest cover transition probabilities between 1993 and 2000, calculated on the basis of the frequency distribution of the observations. The diagonal of the transition probability matrix represents the self-replacement probabilities, i.e. the probability of a forest cover class remaining the same (shown in bold in Table 8.1), whereas the off-diagonal values indicate the probability of a change occurring from one forest cover class to another.

8.2.2 Computation of Transition Potential Maps Using GIS-Based Techniques

Transition potential (suitability) maps represent the likelihood or probability that the landscape will change from one land use/cover class to another (e.g. forest to non-forest). The basic prerequisite for computing transition potential maps is the derivation of weights representing the relative importance of each factor in relationship to a given land use/cover change (Hosseinali and Alesheikh 2008). Generally, weighting methods are classified into data-driven (e.g. WofE, logistic regression and artificial neural networks) and knowledge-driven (analytical hierarchy processes, ratio estimation etc.) groups (Hosseinali and Alesheikh 2008; Liu and Mason 2009). Although both data-driven and knowledge-driven methods have been used for computing transition potential maps (Eastman et al. 2005), the former has the advantage of reducing the problems of biased or incorrect decisions that knowledge-driven methods have (Hosseinali and Alesheikh 2008). While there are many data- and knowledge-driven techniques (Liu and Mason 2009), in this chapter we limit our discussion to the WofE and MCE techniques.

The WofE algorithm uses Bayes' theorem of conditional probability to compute transition potential maps based on the statistical relationship between each land use/cover change (e.g. forest to non-forest) and predictors (i.e. independent variables) such as distance to roads, soil and elevation. This method employs prior and posterior probabilities. The prior probability is defined as the probability of occurrence of a specific land use/cover change, which is calculated by dividing the number of samples (of that land use/cover change) with the total land use/cover change in the study area. The posterior probability is the conditional probability of the existence of a specific land use/cover change (e.g. forest to non-forest) when a predictor variable exists. For example, the conditional probability of the change from forest to non-forest given the presence of predictor variables such as distance to roads, soil and elevation can be expressed as (Almeida et al. 2005; Levine and Block 2011).

$$P\{R|S\} = P\{S|R\} * P\{R\} / P\{S\} \tag{8.3}$$

where $P\{R|S\}$ is the conditional (posterior) probability of the change from forest to non-forest given the presence of the predictor variables, $P\{S|R\}$ is a likelihood function that gives the probability that the predictor variables (data) would be obtained given that R is true, $P\{R\}$ is the prior probability, $P\{S\}$ is the marginal probability of the predictor variables (that is the probability of obtaining the predictor variables under all possible scenarios), R is the change from forest to non-forest, and S is the predictor variables (e.g. distance to roads, soil and slope).

The advantage of WofE is its simplicity and straightforward interpretation of weights (Agterberg and Cheng 2002). However, the basic assumption of the WofE is that predictor variables should be independent. Therefore, the predictor variables should be tested for independence using methods such as the Crammer coefficient (Bonham-Carter et al. 1988; Bonham-Carter 1994; Agterberg and Cheng 2002). Generally, a predictor variable with a Crammer coefficient of more than 0.5 should be removed since it would be highly correlated with other variables (Bonham-Carter et al. 1988; Bonham-Carter 1994).

MCE is a technique for combining data according to its importance in making a decision (Liu and Mason 2009). Many researchers have integrated MCE and GIS (Carver 1991; Jankowski and Richard 1994; Jankowski 1995; Eastman et al. 1995; Wu and Webster 1998). Conceptually, the MCE technique involves qualitative or quantitative weighting, scoring or ranking of criteria to reflect their importance to either single or multiple sets of objectives (Eastman et al. 1995). In essence, the MCE technique uses numerical algorithms that define the "suitability" of a particular solution on the basis of the input criteria and weights, together with some mathematical and logical means of determining trade-offs when conflicts arise (Heywood et al. 1998). Two of the most common procedures for MCE are weighted linear combinations and concordance–discordance analysis (Voogd 1993; Carver 1991). In the former, each factor is multiplied by a weight and then summed to arrive at a final transition potential index. In the latter, each pair of alternatives is analyzed for the degree to which one out-ranks the other in the specified criteria (Eastman et al. 1995). The concordance–discordance analysis is computationally impractical when a large number of alternatives is present (e.g. raster data where every pixel is an alternative), while the weighted linear combination is very straightforward in a raster GIS.

The weighted linear combination (Voogd 1993) combines factors by applying a weight to each factor, followed by a summation of the results to yield a transition potential (suitability) map, i.e.

$$S = SUM\left(w_i * x_i\right) \tag{8.4}$$

where S is suitability (transition potential), w_i is the weight of factor i, and x_i is the criterion score of factor i.

In a case where constraints apply, the procedure can be modified by multiplying the suitability calculated from the factors by the product of the constraints, i.e.

$$S = SUM\left(w_i * x_i\right) * IIc_j \tag{8.5}$$

Where c_j is the criterion score of constraint j, and II is the product.

Most GIS software, such as IDRISI, provides a MCE module developed to compute transition potential maps (Eastman et al. 1995). The primary issues in the computation of transition potential (suitability) maps are the standardization of criteria scores and the development of the factor weights using methods such as the analytic hierarchy process (Saaty 1977; Saaty and Vargas 2001).

8.2.3 Cellular Automata (CA) Models

Cellular automata (CA) are bottom-up, individual-based dynamic models that were originally conceptualized by Ulam and von Neumann in the 1940s in order to understand the behavior of complex systems (Moreno et al. 2010). The CA model consists of an array of cells wherein each cell can assume one of *i* discrete states at any one time (Tobler 1979; Couclelis 1985; White and Engelen 1997). Time progresses in discrete steps, and all cells change their state simultaneously as a function of their own state, together with the state of the cells in their neighborhood, in accordance with a specified set of transition rules (Engelen et al. 1995). In essence, CA encompasses five major components (Wolfram 1984; White and Engelen 1997):

1. A space composed of a regular grid in one or two dimensions
2. A finite set of possible states associated with every cell (e.g. forest or non-forest)
3. A neighborhood composed of adjacent cells whose states influence the central cell
4. Transition rules applied uniformly through time and space
5. A discrete time at which the state of the system is updated

According to Moreno et al. (2010), circular and extended neighborhoods are commonly used to reduce directional bias and capture the spatial influence of surrounding cells on the central one. Space is typically represented as a grid of regular cells, while the neighborhood is defined as a collection of cells based on physical adjacency (White and Engelen 1997; Moreno et al. 2010). Distance functions are applied within a neighborhood to take into account the spatial-dependent attractiveness or repulsiveness of one cell state over another (Soares-Filho et al. 2002; Moreno et al. 2010). In addition to deterministic transition rules, stochastic rules are commonly applied to capture the intrinsic variability of natural and human systems (Moreno et al. 2010). The CA model works by simulating the present based on an extrapolation of the past land use/cover maps. This allows the model to iterate to any other selected date (Messina and Walsh 2001).

For land use/cover changes and urban growth studies, CA models have been found to be more effective than the conventional modeling approaches for a number of reasons. First, CA models allow the integration of macro-scale with micro-scale temporal processes as well as the integration of macro-spatial and micro-spatial phenomena (Wolfram 1984; Torrens 2000). As a result, CA models can make the maximum possible use of available spatial and temporal detail, in contrast to conventional approaches, which operate either at the macro- or the micro-level (Briassoulis 2000). Furthermore, CA models offer a flexible platform for the interaction of biophysical and socio-economic driving factors, as well as for the simulation of real-world complex systems based on simple rules (Wolfram 1984; Engelen 1988; White and Engelen 1997). More importantly, theoretical assumptions may be tested and validated in a particular environmental and socio-economic context (Briassoulis 2000; Torrens 2000). While CA models have produced important contributions to modeling, recent studies have revealed that raster-based CA are sensitive to the modifiable spatial units used in the model, and the modeling results vary according to the cell size and neighborhood configuration (Moreno et al. 2010).

8.2.4 How the MCA Model Works

The spatially explicit nature of the CA model and its compatibility with GIS and other modeling frameworks such as Markov chains have resulted in the development of various hybrid CA models (Walsh et al. 2006). This chapter focuses only on the mechanism of the MCA model based on the Dinamica EGO (environment for geo-processing objects) platform. Dinamica EGO was developed by the Center for Remote Sensing of the Federal University of Minas Gerais in Brazil (Maeda et al. 2011). This MCA model employs an *expander* transition function to expand or contract previous land use/cover class patches, while the *patcher* transition function is used to form new patches through a stochastic seeding mechanism (Soares-Filho et al. 2002). Thus, on the one hand the *expander* transition function performs transitions from state *i* to state *j* only in the neighboring cells of state *j*. On the other hand the *patcher* transition function performs transitions from state *i* to state *j* only in the neighboring cells of states other than *j* (Almeida et al. 2003). First, the algorithm scans the initial land use/cover map to sort out the cells with the highest probabilities and then arrange them in a data array (Almeida et al. 2005). Then the cells are selected randomly from top to bottom of the data array. Finally, the land use/cover map is again scanned to perform the selected transitions (Soares-Filho et al. 2002). If the *expander* transition function does not perform the amount of desired transitions after a fixed number of iterations, it then transfers to the *patcher* transition function a residual number of transitions, so that the total number of transitions always amounts to a desired value (Soares-Filho et al. 2002). The desired transitions are obtained from Markov chain-computed transition probabilities. However, the *patcher* transition function will simulate land use/cover change patterns by generating diffused patches, while at the same time preventing the formation of single isolated one-cell patches (Almeida et al. 2003). This function searches for cells around a chosen location for a given transition through the selection of the core cell of the new patch based on a specific number of cells around the core cell according to their transition probabilities (Soares-Filho et al. 2002).

The *expander* and *patcher* transition functions are composed of an allocation mechanism responsible for identifying cells with the highest transition probabilities for each *ij* transition. As a result, cells are stored and organized for later selection. The two complementary functions (i.e. the *expander* and the *patcher*) consist of mean patch size, patch size variance and isometry parameters, which can be changed to produce various spatial patterns of land use/cover patches according to a log–normal probability distribution function (Soares-Filho et al. 2002). For example, an increase in mean patch size results in a less fragmented landscape, while an increase in the patch size variance results in a more diverse landscape (UFMG 2009). Isometry is a number that varies from 0 to 2, and thus an isometry greater than one results in more isometric (equal) patches (UFMG 2009). Finally, MCA model iterations are specified according to time differences between two land use/cover maps ($\Delta t = t_2 - t_1$).

8.3 Application of MCA Models in Previous Studies

Modeling approaches that integrate CA and Markov chains have been explored for some time (Zhou and Liebhold 1995; Li and Reynolds 1997; Parker et al. 2003; Aspinall 1994). A major advantage of the MCA approach is that GIS and remote sensing data can be incorporated effectively (Li and Reynolds 1997). In particular, biophysical and socio-economic data can be used to define initial conditions, to parameterize the MCA model, to calculate transition probabilities and to determine the neighborhood rules with transition potential maps.

Although the potential of the MCA models has been recognized, few studies have used the MCA models for simulating land use/cover changes. Li and Reynolds (1997) developed a combined Markov and CA model to simulate the effects of spatial pattern, drought and grazing on the rates of rangeland degradation. Although their model was conceptually appealing, it did not account for the variations of transition probabilities due to changes in environmental, socio-economic and political factors. To overcome such limitations, Soares-Filho et al. (2002) incorporated a *saturation value* parameter that is designed to vary the transition rates through a dynamic feedback analysis of landscape changes. Their spatially explicit, multi-scale and dynamic stochastic CA modeling framework successfully simulated land use/cover changes in the Amazonian colonization frontier (Soares-Filho et al. 2002; Soares-Filho et al. 2006). Recently, the Dinamica EGO modeling framework has also introduced a scenario generator model that computes transition rates based on the integration of environmental and socio-economic factors (Almeida et al. 2005; Teixerira et al. 2009).

Pontius and Malanson (2005) applied the MCA to predict land use/cover changes in central Massachusetts. Their model used an MCE technique to compute transition potential maps, and a spatial contiguity rule to determine the location of predicted change. Contemporary legal constraint data were used as an additional driver to calibrate the transition potential (suitability) maps (Pontius and Malanson 2005). Although the MCA model produced good results, it did not incorporate additional constraints and factors that represent socio-economic and urban planning issues. Furthermore, the authors concluded that their MCA model was poor at predicting the location of built to non-built conversions (Pontius and Malanson 2005). Paegelow and Olmedo (2005) also used the MCA model for testing the possibilities and limits of a prospective land cover modeling in France and Spain. Their model used the Markov chain analysis to control temporal dynamics, while MCE, multi-objective evaluation and CA controlled spatial contiguity in order to determine the location of the predicted land cover change. Land cover maps and relevant environmental factors were used to calibrate the transition potential (suitability) maps. While the authors reported an overall accuracy of 75%, they noted the need to analyze prediction residues in order to improve the model.

Myint and Wang (2006) also applied the MCA for projecting land use/cover changes in Norman, Oklahoma, USA. Their model also used an MCE technique to compute transition potential maps, and a spatial contiguity rule to determine the location of predicted change. Ancillary map layers such as roads and drainage were used as driving factors in order to calibrate the transition potential (suitability) maps. The suitability ratings were based on the authors' personal judgment in consultation with land use planners, which may possibly lead to bias (Hosseinali and Alesheikh 2008). Although their model was effective at projecting future land use/cover changes, as indicated by an overall accuracy of 86.2%, their accuracy assessment procedure only considered accuracy in terms of quantity and not in terms of location (Pontius and Malanson 2005). More recently, Kamusoko et al. (2009) applied a MCA model in rural areas in Zimbabwe. Their model's overall simulation success was 69% for the 2000 simulated land use/cover map, and 83% for the 2005 simulated land use/cover map. However, the authors reported that the model was poor at simulating the location of bare land areas owing to the lack of input spatial data.

8.4 Current Status and Future Prospects

The increasing awareness of the impact of land use/cover changes on global climate change has renewed interest in the application of spatial simulation models (Soares-Filho et al. 2006; Brown et al. 2007). For example, initiatives that are currently being negotiated under the United Nations Framework Convention on Climate Change (UNFCCC) to reduce emissions from deforestation and forest degradation in developing countries requires the development of robust baseline or reference scenarios under the business-as-usual (BAU) scenario (Angelsen et al. 2009). A baseline or reference scenario (under BAU) is the projected deforestation and associated emissions in the absence of a REDD (reducing emissions from deforestation and forest degradation) project (Angelsen 2008). Several approaches for setting baseline or reference scenarios have been suggested, which include among others spatial and non-spatial modeling approaches (Brown et al. 2007; Terrestrial Carbon Group 2008). However, this new interest in spatial simulation models also presents new challenges to researchers and decision makers because the establishment of robust baseline or reference scenarios requires a better understanding of the underlying driving forces in order to capture intrinsic landscape processes at multiple spatial and temporal scales (GOFC-GOLD 2010). In addition, attention should also be focused on new theoretical and methodological developments in the modeling framework. This section highlights the current status and future prospects of MCA models, paying special attention to issues pertaining to (1) theories underpinning model development, (2) data issues, and (3) calibration and validation.

8.4.1 Theories Underpinning Model Development

Current spatial simulation models of land use/cover changes can be broadly divided into those which are based on theory and those which are not (Verburg et al. 2004). The former include mainly economic theory-based models as well as spatial interaction models (Lambin et al. 2000; Irwin and Geoghegan 2001; Verburg et al. 2004;

Soares-Filho et al. 2006), while the latter comprise models that do not include theory explicitly, or those that are based on specific theoretical assumptions (Myint and Wang 2006; Kamusoko et al. 2009). Although theory is critical during model specification and interpretation, the influence of theory and assumptions on the modeling results is not always examined (Verburg et al. 2004). This unfortunately limits the reliability and robustness of the model (Briassoulis 2000). To overcome this limitation, future MCA models will need to incorporate a strong theoretical background which is relevant to the given underlying landscape processes. This requires more collective efforts that focus on developing an integrated and multidisciplinary research paradigm. The land use/cover change modeling community has been working on a number of multi-disciplinary research programs aimed at improving spatial models (Geoghegan et al. 1998; Irwin and Geoghegan 2001).

8.4.2 Data and Scale Issues

Fundamental to the development of robust MCA models are issues such as the spatial and temporal dimensions, reliability, availability and cost of data collection (Briassoulis 2000). In most cases, the spatial units usually follow administrative boundaries, which, although appropriate for policy implementation, may not be meaningful for all types of data (Verburg et al. 2004). With respect to the temporal dimension, the temporal systems of reference (e.g. time and number of observations) are not always compatible and consistent (Verburg et al. 2004). In other words, different definitions among time periods, especially at lower levels of aggregation, give rise to problems of compatibility and consistency, particularly with historical data. For example, the dates of historical land use/cover maps may not be compatible with the available socio-economic data, which may have been acquired at a different time. Furthermore, MCA models are built on the assumptions of temporal homogeneity and progressive linear trends, despite the fact that land use/cover changes have occurred in the context of long-term instability (e.g. fluctuations in climate, prices or state policies). These issues are important for land use/cover models where the exact time and length of the policy intervention is critical in the modeling framework. In addition, the availability and cost of obtaining proper longitudinal data (e.g. socio-economic data) limit the reliability of land use/cover change models that integrate biophysical and socio-economic data.

With reference to the spatial dimension, past studies have revealed that raster-based CA models are sensitive to the modifiable spatial units used in the model, and that results vary according to the cell size and neighborhood configuration (Veldkamp et al. 2001; Chen and Mynett 2003; Jantz and Goetz 2005). To overcome the sensitivity of raster-based CA models to cell size and neighborhood configurations, novel geographic objected-based CA models have been developed (Torrens and Benenson 2005; Moreno et al. 2010). According to Moreno et al. (2010), space is defined as a collection of geographic objects of irregular shape and size corresponding to meaningful real-world features. Furthermore, the neighborhood is dynamic (i.e. it includes the whole geographic space), and the model allows the geometric transformation of each object according to a transition function that incorporates the influence of its neighbors (Moreno et al. 2010).

8.4.3 Calibration and Validation

Calibration and validation are important components in the development of MCA models. However, validation is the weakest part of land use/cover modeling, since there are no agreed criteria to assess the performance of one land use/cover model versus another, or to compare one run versus another run of the same model (Pontius et al. 2004). In order to assess the model's predictive power, a clear distinction between the procedures for calibration and validation must be made, the failure of which makes the interpretation of any results difficult or misleading (Pontius et al. 2004). In some cases, it is more common to force the prediction to simulate the correct quantity of each land use/cover class, than to assess whether the model predicts the correct location of land use/cover (Kok et al. 2001; Pontius et al. 2001). Any lack of clarity in the methodology to distinguish the calibration information from the validation information causes confusion in land use/cover modeling, which can lead to a misunderstanding of the model's certainty (Pontius et al. 2004).

Calibration is "the estimation and adjustment of model parameters and constraints to improve the agreement between model output and a data set" (Rykiel 1996). The information used for calibration should be at or before some specific point in time (t1), which is the point in time at which the predictive extrapolation begins. In contrast, validation is the process of comparing the model's prediction for t2 with a reference map of time t2, where the reference map is considered to be a much more accurate portrayal of the landscape at time t2 (Pontius and Malanson 2005). One set of data should be used to calibrate the model, and a separate set should be used to validate the model (Pontius et al. 2004). In order to enhance the validity of land use/cover modeling, Pontius et al. (2004) suggested that it is helpful to use a validation technique that, (a) takes into account the source of error, (b) compares the model to a null model (a model that predicts pure persistence, i.e. no change between t1 and t2), and (c) performs analysis at multiple scales.

8.5 Summary and Conclusions

This chapter has attempted to review the current state-of-the-art operational MCA land use/cover change models. Despite the existence of the many land use/cover change modeling challenges highlighted in this chapter, the land use/cover modeling research community has developed a variety of models, which have been applied with varying success in different regions of the world. Interesting data sets, as well as the functioning of interdisciplinary and multi-disciplinary research teams, have made efforts to improve and develop robust land use/cover change models that can be useful for understanding the functioning of land use/cover systems, and also to support land use planning and policy (Lambin et al. 2000; Irwin and Geoghegan 2001; Rindfuss et al. 2003; Verburg et al. 2004). However, the review has also exposed the limitations of the current MCA models and modeling practice. Many current MCA land use/cover models are still built on common assumptions of homogeneity and linear trends, which may fail to capture the underlying real-world landscape processes characterized by non-linear trends. While much effort has been spent on model calibration, little attention has been given to the development of robust validation methods (Pontius and Malanson 2005).

Nonetheless, the limitations singled out present opportunities for research into MCA models. Future MCA land use/cover change models will need to be more integrated and more responsive to different environmental, socio-economic and political conditions (Verburg et al. 2004). Given the rapid developments in computer technology (increases in memory and speed of computers), more integrated MCA models should be developed. For example, encouraging research is being done in the area of geographic object-based CA models (Torrens and Benenson 2005; Moreno et al. 2010). These novel geographic object-based CA models should be incorporated in the MCA modeling framework. More research should also be done in developing non-linear Markov chains that can compute non-linear transition probabilities. In addition, research will also have to address the problems of the evaluation of policy impacts as well as issues of household decision-making. Predominantly, aggregate modeling techniques need to be complemented by agent-based methods capable of measuring the influence of individuals and communities on land use/cover changes. The feasibility of such research would be greatly enhanced by the availability of the detailed land use/cover, biophysical and disaggregate socio-economic data required for integrated agent-based MCA models (Berger 2001).

Finally, more efforts should be made to disseminate land use/cover models in general, and MCA models in particular, by including institutions and individuals, particularly in developing countries. This must be supported by the development of user-friendly modeling software such as Dinamica EGO (Soares-Filho et al. 2002) and IDRISI Taiga (Eastman 2009). Although, spatial simulation models have been criticized for failing to adapt to new challenges and problems, researchers and decision makers are collaborating in order to develop robust MCA land use/cover change models (Rindfuss et al. 2003). These models would be useful for understanding the driving forces and underlying processes of land use/cover changes, as well as to simulate future land use/cover changes.

References

Agterberg FP, Cheng Q (2002) Conditional independence test for weights-of-evidence modeling. Nat Resour Res 11(4):249–255

Almeida CM, Batty M, Monteiro AMV, Camara G, Soares-Filho BS, Cerqueira GC et al (2003) Stochastic cellular automata modeling of urban land use dyanimics: empirical development and estimation. Comput Environ Urban Syst 27:481–509

Almeida CM, Monteiro AMV, Camara G, Soares-Filho BS, Cerqueira GC, Pennachin CL, Batty M (2005) GIS and remote sensing as tools for the simulation of urban land-use change. Int J Rem Sens 26(4):759–774

Angelsen A (2008) REDD models and baselines. Int For Rev 10(3):465–475

Angelsen A, Brockhaus M, Kanninen M, Sills E, Sunderlin WD, Wertz-Kanounnikoff S (eds) (2009) Realising REDD+: national strategy and policy options. CIFOR, Bogor

Aspinall R (1994) Use of GIS for interpreting land-use policy and modeling effects of land use change. In: Haines-Young R, Green DR, Cousins S (eds) Landscape ecology and geographic information systems. Taylor and Francis, London, pp 223–236

Batty M (1998) Urban evolution on the desktop: simulation with the use of extended cellular automata. Environ Plann Plann Des 30:1943–1967

Bell EJ (1975) Stochastic analysis of urban development. Environ Plann 7:35–39

Bell EJ, Hinojosa RC (1977) Markov analysis of land use change: continuous time and stationary processes. Soc Econ Plann Sci 11:13–17

Berger T (2001) Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion. Resource use changes and policy analysis. Agr Econ 25:245–260

Boerner REJ, DeMers MN, Simpson JW, Artigas FJ, Silva A, Berns LA (1996) Markov models of inertia and dynamic on two contiguous Ohio landscapes. Geogr Anal 28:56–66

Bonham-Carter G (1994) Geographic information systems for geoscientists: modelling with GIS. Pergamon, New York

Bonham-Carter GF, Agterberg FP, Wright DF (1988) Integration of geological data sets for gold exploration in Nova Scotia. American Society for Photogrammetry and Remote Sensing, Maryland

Briassoulis H (2000) Analysis of land use change: theoretical and modeling approaches. Regional Research Institute, West Virginia University, Morgantown. http://www.rri.wvu.edu/WebBook/Briassoulis/contents.html. Accessed 14 May 2005

Brown S, Hall M, Andrasko K, Ruiz F, Marzoli W, Guerrero G, Masera O, Dushku A, De Jong B, Cornell J (2007) Baselines for land-use change in the tropics: application to avoided deforestation projects. Mitigation and Adaptation Strategies for Global Change 12:1001–1026

Carver S (1991) Integrating multi-criteria evaluation with geographical information systems. Int J Geogr Inform Syst 5:321–339

Chen Q, Mynett AE (2003) Effects of cell size and configuration in cellular automata based prey-predator modeling. Simul Model Pract Theory 11(7–8):609–625

Cheng J, Masser I (2004) Understanding spatial and temporal processes of urban growth: cellular automata modelling. Environ Plann Plann Des 31:167–194

Clarke KC, Hoppen S, Gaydos L (1997) A Self-modifying cellular automaton model of historical urbanization in the San Fransisco Bay Area. Environ Plann 24:247–261

Couclelis H (1985) Cellular worlds: a framework for micro–macro dynamics. Environ Plann 17:585–596

Couclelis H (1989) Macrostructure and microbehavior in a metropolitan area. Environ Plann 16:141–154

Dale VH (1997) The relationship between land-use change and climate change. Ecol Appl 17(3):753–769

UFMG (Universidade Federal de Minas Gerais) (2009) Dinamica EGO. http://www.csr.ufmg.br/dinamica/. Accessed 2 Apr 2009

Drewett JR (1969) A stochastic model of the land conversion process. Reg Stud 3:269–280

Eastman JR (2009) Idrisi Taiga, guide to GIS and image processing. Clark University Edition, p 342

Eastman JR, Jin W, Kyem PAK, Toledano J (1995) Raster procedures for multi-criteria/multi-objective decisions. Photogramm Eng Rem Sens 61(5):539–547

Eastman JR, Solorzano LA, van Fossen ME (2005) Transition potential modeling for land-cover change. In: Maguire DJ, Batty M, Goodchild MF (eds) GIS, spatial analysis, and modeling. ESRI Press, California, pp 357–385

Engelen G (1988) The theory of self-organization and modeling complex urban systems. Eur J Oper Res 37:42–57
Engelen G, White R, Uljee I, Drazan P (1995) Using cellular automata for integrated modeling of socio-environmental systems. Environ Monit Assess 34:203–214
Etter A, McAlphine C, Wilson K, Phinn S, Possingham H (2006) Regional patterns of agricultural land use and deforestation in Colombia. Agric Ecosyst Environ 114:369–386
Gar-On Yeh A, Li X (2009) Cellular automata, and GIS for urban planning. In: Madden M (ed) Manual of geographic information systems. ASPRS, Sacramento, pp 591–619
Geoghegan J, Pritchard L Jr, Ogneva-Himmelberger Y, Chowdhury RR, Sanderson S, Turner BL II (1998) Socializing the pixel and pixelizing the social in land-use and land-cover change. In: Liverman D, Moran EF, Rindfuss RR, Stern PC (eds) People and pixels: linking remote sensing and social science. National Academy Press, Washington, pp 151–169
GOFC-GOLD (2010) A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals caused by deforestation, gains and losses of carbon stocks in forests, remaining forests, and forestation. GOFC-GOLD Report version COP16-1. GOFC-GOLD Project Office, Natural Resources Canada, Alberta
Heywood I, Cornelius S, Carver S (1998) An introduction to geographical information systems. Addison Wesley Longman, England, pp 138–142
Hosseinali F, Alesheikh AA (2008) Weighting spatial information in GIS for copper mining exploration. Am J Appl Sci 5(9):1187–1198
Irwin EG, Geoghegan J (2001) Theory, data, methods: developing spatially explicit economic models of land use change. Agric Ecosyst Environ 85:7–23
Jahan S (1986) The determination of stability and similarity of Markovian land use change processes: a theoretical and empirical analysis. Soc Econ Plann Sci 20:243–251
Jankowski P (1995) Integrating geographical information systems and multiple criteria decision making methods. Int J Geogr Inf Syst 9(3):251–273
Jankowski P, Richard L (1994) Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. Environ Plann 21:399–420
Jantz CA, Goetz SJ (2005) Analysis of scale dependencies in an urban land-use change in the central Arizona–Phoenix region, USA. Landsc Ecol 16:611–626
Kaimowitz D, Angelsen A (1998) Economic models of tropical deforestation: a review. Center for International Forestry Research, Bogor
Kamusoko C, Aniya M, Bongo A, Munyaradzi M (2009) Rural sustainability under threat in Zimbabwe – simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Appl Geogr 29(3):435–447
Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic, New York
Kok K, Farrow A, Veldkamp TA, Verbug PH (2001) A method and application of multi-scale validation in spatial land use models. Agr Ecosyst Environ 85(1–3):223–238
Lambin EF (1994) Modelling deforestation processes: a review. TREES Publications Series B: Research Report no.1, European Commission, EUR 15744 EN
Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. Prog Phys Geogr 21:375–393
Lambin EF, Rounsevell M, Geist H (2000) Are agricultural land-use models able to predict changes in land use intensity? Agr Ecosyst Environ 82(1–3):321–331
Levine N, Block R (2011) Are Bayesian journey-to-crime estimation: an improvement in geographic profiling methodology. Prof Geogr 63(2):213–229
Li H, Reynolds JF (1997) Modeling effects of spatial pattern, drought, and grazing on rates of rangeland degradation: a combined Markov and cellular automaton approach. In: Quattrochi DA, Goodchild MF (eds) Scale in remote sensing and GIS. Lewis Publishers, Boca Raton, pp 211–230
Liu JG, Mason PJ (2009) Essential image processing and GIS for remote sensing. Wiley-Blackwell, Oxford
Maeda EE, Almeida CM, Ximenes AC, Formaggio AR, Shimabukuro YE, Pellikka P (2011) Dynamic modeling of forest conversion: Simulation of past and future scenarios of rural

activities expansion in the fringes of the Xingu National Park, Brazilian Amazon. Int J Appl Earth Observation Geoinformation 13:435–446

Matheussen B, Kirschbaum RL, Goodman IA, O'Dennel GM, Lettenmaier DP (2000) Effects of land cover change on stream flow in the interior Columbia river basin (USA and Canada). Hydrolog Process 14(5):867–885

Messina J, Walsh S (2001) 2.5D morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. Plant Ecol 156:75–88

Moreno N, Quintero R, Ablan F, Barros F, Davila J, Ramirez H, Tonella G, Acevedo MF (2007) Biocomplexity of deforestation in the Caparo tropical forest reserve in Venezuela: an integrated multi-agent and cellular automata model. Environ Model Softw 22:664–673

Moreno N, Wang F, Marceau DJ (2010) A geographic object-based approach in cellular automata modeling. Photogramm Eng Rem Sens 76(2):183–191

Muller MR, Middleton J (1994) A Markov model of land use change dynamics in the Niagara Region, Ontario, Canada. Landsc Ecol 9:151–157

Myint SW, Wang L (2006) Multicriteria decision approach for land use land cover change using Markov chain analysis and a cellular automata approach. Can J Remote Sens 32(6):390–404

Paegelow M, Olmedo MTC (2005) Possibilities and limits of prospective GIS land cover modelling: a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). Int J Geogr Inform Sci 19(6):697–722

Parker DC, Manson SM, Janssen MA, Hoffman M, Deadman P (2003) Multi-agent systems for the simulation of land-use and land-cover change: a review. Ann Assoc Am Geogr 93(2):314–337

Pastor J, Bonde J, Johnston C, Naiman RJ (1993) Markovian analysis of the spatially dependent dynamics of beaver ponds. In: Gardner RH (ed) Predicting spatial effects in ecological systems: lectures on mathematics in the life sciences, 23. American Mathematical Society, Providence, pp 5–27

Petit C, Scudder T, Lambin E (2001) Quantifying processes of land-cover change by remote sensing: resettlement and rapid land-cover changes in south-eastern Zambia. Int J Rem Sens 22:3435–3456

Pijanowski BC, Brown DG, Shellito BA, Manik GA (2002) Using neural networks and GIS to forecast land use changes: a land Transformation Model. Comput Environ Urban Syst 26:553–575

Pontius RG Jr, Malanson J (2005) Comparison of the structure and accuracy of two land change models. Int J Geogr Inf Sci 19(2):243–265

Pontius RG Jr, Cornell JD, Hall CAS (2001) Modelling spatial patternsof land use change with GEOMOD: application and validation for Costa Rica. Agr Ecosyst Environ 85(1–3):553–575

Pontius RG Jr, Huffaker D, Denman K (2004) Useful techniques of validation for spatially explicit land-change models. Ecol Model 179:445–461

Rindfuss RR, Walsh SJ, Mishra V, Dolcemascalo GP (2003) Linking household and remotely sensed data: methodological and practical problems. In: Fox J, Rindfuss RR, Walsh SJ, Mishra V (eds) People and the environment: approaches for linking household and community surveys to remote sensing and GIS. Kluwer Academic, Boston, pp 1–29

Robinson VB (1978) Information theory and sequences of land use: an application. Prof Geogr 30:174–179

Rykiel EJ Jr (1996) Testing ecological models: the meaning of validation. Ecol Model 90:229–244

Saaty TL (1977) A scaling method for priorities in hierarchical structures. J Math Psychol 15:234–281

Saaty TL, Vargas LG (2001) Models, methods, concepts and application of analytic hierarchy process. Kluwer Academic, Boston

Silverton J, Holtier S, Johnson J, Dale P (1992) Cellular automaton models of interspecific competition for space – the effect of pattern on process. J Ecol 80:527–534

Soares-Filho BS, Cerqueira GC, Pennachin CL (2002) DINAMICA: a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecol Model 154:217–235

Soares-Filho BS, Nepsta DC, Curran LM, Cerqueira GC, Garcia RA, Ramos CA, Voll E, McDonald A, Lefebvre P, Schlesinger P (2006) Modelling conservation in the Amazon basin. Nature 23:520–523

Teixerira AMG, Soares-Filho BS, Freitas SR, Metger JP (2009) Modeling landscape dynamics in an Atlantic rainforest region: implications for conservation. For Ecol Manag 257:1219–1230

Terrestrial Carbon Group (2008) How to include terrestrial carbon in developing nations in the overall climate change solution. http://www.terrestrialcarbon.org/. Accessed 14 Jan 2011

Tobler W (1979) Cellular geography. *In:* Gale S and Olsson G (eds) Philosophy in geography. Reidel Publishing Company, Dortrecht, pp 379–386

Torrens PM (2000) How cellular models of urban systems work (1. Theory). CASA Working Paper Series*, 28*. Centre for Advanced Spatial Analysis (UCL), London. http://www.casa.ucl.ac.uk/working_papers/paper28.pdf. Accessed 17 Aug 2009

Torrens PM (2006) Simulating sprawl. Ann Assoc Am Geogr 96(2):248–275

Torrens PM, Benenson I (2005) Geographic automata systems. Int J Geogr Inform Sci 19(4):385–412

Veldkamp A, Verbug PH, Kok K, de Koning GHJ, Priess J, Bergsma AR (2001) The need for scale sensitive approaches in spatially explicit land use change modelling. Environ Model Assess 6:111–121

Verburg PH, Schot PP, Dijst MJ, Veldkamp A (2004) Land use change modelling: current practice and research priorities. GeoJournal 61:309–324

Voogd H (1993) Multicriteria evaluation for urban and regional planning. Pion, London

Walsh SJ, Entwilse B, Rindfuss RR, Pages PH (2006) Spatial simulation modelling land use/land cover change scenarios in northern Thailand: a cellular automata approach. J Land Use Sci 1(1):5–28

Weng Q (2002) Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. J Environ Manag 64:273–284

White R, Engelen G (1997) Cellular automata as the basis of integrated dynamic regional modeling. Environ Plann 24:235–246

Wolfram S (1984) Cellular automata as models of complexity. Nature 311:419–424

Wood EC, Tappan GG, Hadj A (2004) Understanding the drivers of agricultural land use change in south-central Senegal. J Arid Environ 59:565–582

Wu F, Webster CJ (1998) Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environ Plann 25:103–126

Wu Q, Li H, Wang R, Paulusen J, He Y, Wang M, Wang Z (2006) Monitoring and predicting land use changes in Beijing using remote sensing and GIS. Landsc Urban Plann 78(4):322–333

Zhou G, Liebhold AM (1995) Forecasting the spread of gypsy moth outbreaks using cellular transition models. Landsc Ecol 10:177–186

Chapter 9
Multi-layer Perceptron Neural Networks in Geospatial Analysis

Duong Dang Khoi and Yuji Murayama

9.1 Introduction

Geospatial analysis involves using a variety of approaches. Deciding on a suitable approach depends on the complexity of the problem being addressed and the degree to which the problem is understood. Several algebraic and numerical computing techniques can be used to describe the behavior and nature of real geographical processes and develop mathematical models to represent them. Such methods require accurate knowledge of the process dynamics to emulate the processes. However, in practice, the knowledge required to solve a problem may be incomplete because the source of the knowledge is unknown, or because the complexity of the problem may introduce uncertainties and inaccuracies that make modeling unrealistic. In this case, an approximate analysis approach can be used. Artificial neural networks (ANNs) are an open concept that allows for the continuous refinement and acquisition of new knowledge and can provide solutions to such problems. It offers an alternative for dealing with solutions with a tolerance of imprecision, uncertainty, and approximation. It is known as an information-processing paradigm inspired by the interconnected and parallel structure of the human brain. The initial concepts of ANNs were attempts to depict the characteristics of biological neural networks in order to address a series of information-processing issues. This research

D.D. Khoi (✉)
Faculty of Land Administration, Hanoi University of Natural Resources and Environment, Vietnam

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: khoi_tn@yahoo.com

Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: mura1@sakura.cc.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*, DOI 10.1007/978-4-431-54000-7_9, © Springer Japan 2012

domain has been extensively studied and applied during the last three decades. ANNs provide a flexible data analysis framework for appropriate nonlinear mappings from a variety of input variables. In particular, they are distribution-free and thus have an advantage over most statistical methods that require knowledge of the distribution function. In particular, they can learn complex functional relationships between input and output data that are not envisioned by researchers (Kim and Nelson 1998). ANNs are applied in a wide range of fields, such as medicine, molecular biology, ecology, environmental sciences, and image classification (Atkinson and Tatnall 1997).

A variety of ANN types have been developed to capture particular features of the neurons and their interaction to solve different kinds of application problems. Each ANN model has a typical training algorithm. Some ANN models are suitable for theoretical investigations. Much of the modern effort in the modeling of nervous systems has been for pattern recognition tasks. The most commonly used networks are multilayer perceptron (MLP) networks, radial basis function (RBF) networks, recurrent neural networks (RNNs), and Kohonen self-organizing map (SOM) networks.

MLP networks are general, flexible, and nonlinear models with a number of units organized into multiple layers which map the inputs to the desired outputs by minimizing the errors between the desired outputs and the calculated outputs. The complexity of the network can be adjusted by varying the number of layers and the number of units in each layer. Given enough hidden units and enough data, it has been shown that MLPs can approximate any function to any desired accuracy. MLPs can be trained with several different algorithms, but most commonly with the back-propagation training algorithm.

The RBF three-layer architecture is similar to that of MLPs, but the output is computed as a linear combination of basic functions. The hidden layer computes the distance from the input data to each of the centers. These networks have the advantage of being much simpler than the perceptrons while keeping their key nature of a universal approximation of multi-variate functions (Poggio and Girosi 1987). Each node in the hidden layer of the network evaluates a RBF based on incoming inputs. In particular, these networks are introduced with the centers of data clusters; therefore, they have some advantages over the MLP, such as faster convergence, smaller extrapolation errors, and greater reliability (Girosi and Poggio 1990). The linear solutions for the connection weights of the network directly yield a unique local minimum and global minimum. This kind of network significantly shortens the training of the network. The RBF can overcome the possibility of producing complex error surfaces, which are produced by the back-propagation algorithm. RBF networks combine nonlinear RBF activation functions in a hidden layer and the linear combination of weights in the output layer. Different activation functions, e.g., k-means clustering (Hartigan and Wong 1979) or fuzzy c-means (Zhu and He 2006), can be used. However, in most applications of RBF networks, the preferred choice is the Gaussian function.

RNNs are a type of ANN in which connections between nodes form a directed cycle. This creates an internal state of the network that allows the exhibition of dynamic behavior. RNNs are good at characterizing dynamic systems. In this type of network, the hidden node activation values, or network output values, feed back into the network input layer nodes. Internally, a RNN is identical to a MLP. It has been extended from the MLP to include recurrent connections in order to perform temporal processing. An advantage of recurrent networks is that time is represented implicitly in the architecture by incorporating a form of short-term memory implemented through feedback connections. Back-propagation through time (BPTT), modified for multiple recurrent connections, is used as the training algorithm of the RNNs.

The MLP works well only when training data are available, so that the network is able to learn all the patterns that exist in the study area. However, in practice, training data are not always available. Given this weakness, it is critical to develop an appropriate network model to deal with the problem. SOM networks, a neural network type developed by Kohonen (1984, 1995), can be an alternative to the MLP. In the SOM, neighboring pixels in a neural network compete in their activities by means of mutual lateral interaction, and then develop adaptively into different patterns (Kohonen 1990). SOMs are primarily used to visualize and interpret high-dimensional data sets. SOMs can perform the modeling of data sources and provide preferable types of cell neighborhood arrangements, such as planar grid, cylinder, and toroidal. They can be used as a data mining tool, especially for multi-band image classification. SOMs are trained by an unsupervised algorithm by which input data are self-organized into clusters.

The entire spectrum of ANNs is very diverse. In this chapter, attention is limited to one class of ANNs, namely MLP neural networks or feed-forward neural networks. They are becoming an increasingly powerful technique for solving a wide range of problems. Although MLP networks are the most popular and widely used models in many practical applications, many issues and problems (such as network structure, slow convergence speed, and local minima sticking) in applying MLPs have motivated research into improving the standard methods. Much progress has been made in improving the performance of standard MLPs. In particular, additional improvements in training the networks are needed, as the training process is chaotic in nature. The purpose of this chapter is to review the state of the art of the MLP. Specifically, we review the history of MLP development. Then the standard architecture and algorithm of the MLP are described. Finally, MLP applications in land-cover classification and future prospects are summarized.

9.2 History of Perceptron Neural Networks

ANNs are a popular data mining and image processing tool. They originated from attempts to model human thought as an algorithm that can run on a computer. ANNs offer a method of describing artificial neurons to deal with complex problems in the same manner as the human brain. The origin of ANN research was an interest in the study of the mechanism and structure of the human brain. For many years, particularly since the middle of the last century, the mechanism and structure of the brain have been studied in increasing depth. This research has resulted in the development of ANNs.

Animal nervous systems are composed of thousands or millions of interconnected cells. Each of them is a very complex arrangement which deals with incoming signals in many different ways. Our present knowledge of the structure and physiology of neurons is the result of 100 years of research in this field between 1850 and 1950. The transmission of sensory signals (information) is one of the important functions of the neurons. Neurons or nerve cells receive multiple input stimuli, combine and modify the inputs in some way, and then transmit the result to other neurons. Scientific results indicate that neurons can transmit information using electrical signals. Around 1901, Santiago Ramon y Cajal postulated that the specific networking of the nervous cell determines a direction for the transmission of information. The chemical transmission of information at the synapses was studied from 1920 to 1940. The Hodgkin–Huxley model was in some ways one of the first artificial neural models (Cronin 1987), because the postulated dynamics of the nerve impulses could be simulated with simple electrical networks (Mead 1989). The mechanisms for the production and transport of signals from one neuron to another are well-understood physiological phenomena, but how these individual systems cooperate to form complex and massively parallel systems, capable of incredible feats of information processing, has not yet been completely elucidated. Mathematics, physics, and computer science can provide invaluable help in the study of these complex systems. The mathematical properties of ANNs were studied by nonbiological researchers such as Warren McCulloch, Walter Pitts, and John von Neumann. Research in the neurobiological field has progressed in close collaboration with mathematicians and computer scientists.

Neurons receive sensory signals and produce a response. The general structure of a biological neuron consists of dendrites, synapses, a cell body, and an axon. Dendrites are the transmission channels for incoming information. They receive the signals at the contact regions with other cells, called synapses. The body of the cell produces all the necessary chemicals for the continuous working of the neuron. The output signals are transmitted by the axon. Artificial neurons for computing will have input channels, a cell body, and an output channel. Since the middle of the last century, increasing research in the field of artificial intelligence has resulted in the development of computational models of ANNs based on a biological background for solving pattern recognition and data processing tasks. Biological neural networks are just one of many possible solutions to the problem of processing information. Biological neural networks are self-organizing systems, and each individual neuron is a delicate self-organizing structure capable of processing information in many different ways.

The explanation of many important aspects of the physiology of neurons set the stage for the formulation of ANN models. In particular, Hebb (1949) described the adoption laws about the organization of behaviors. These principles also played a significant part in the investigation of the neural foundations of behavior because they provide a general framework for relating behavior to synaptic organization through the dynamics of neural networks. Hebb was the first to examine the mechanisms by which environment and experience can influence brain structure and function, and his ideas formed the basis for work on enriched environments as stimulants for behavioral development. ANNs constitute an alternative computability paradigm. Biological neural networks give us a clue to the properties which would be interesting to include in our artificial networks. ANNs have aroused intense interest in recent years, not only because they exhibit interesting properties, but also because they try to mirror the kind of information processing capabilities of nervous systems. ANNs can be considered as just another approach to the problem of computation.

ANNs emerged after the introduction of a simplified neuron, a mathematical model of the neuron by McCulloch and Pitts (1943). McCulloch and Pitts are generally recognized as the designers of the first neural network, and they attempted to translate the events in the nervous system into rules governing information. This abstract neuron provided the foundations for a formal calculus of brain activity. Rosenblatt (1958), an American psychologist, proposed the "perceptron," a more general computational model than McCulloch and Pitts' units, and this idea subsequently received much attention. The critical innovation was the introduction of numerical weights and a special interconnection pattern. In the Rosenblatt model, computing units are threshold elements and the connectivity is determined stochastically. Learning takes place by adapting the weights of the network with a numerical algorithm. Rosenblatt's model was refined and perfected in the 1960s, and its computational properties were carefully analyzed by Minsky and Papert (1969).

The concept of the perceptron (Rosenblatt 1958) was the starting point for the development of many types of the ANNs. A *perceptron* is a device that computes a weighted sum of its inputs and puts this sum through a special function, known as the activation function, to produce the output. The activation function can be linear or nonlinear. The perceptron is able to classify linearly separable data, but it is unable to handle nonlinear data. The inputs are fed directly to the outputs via a series of weights. In this way it can be considered to be the simplest kind of feed-forward network. The sum of the products of the weights and the inputs is calculated, and if the value is above a certain threshold, the neuron fires and takes the activated value; otherwise it takes the deactivated value. Neurons with this kind of activation function are also called artificial neurons or linear threshold units. Although the perceptron initially seemed promising, it was proved that a single perceptron cannot be trained to recognize many classes of data patterns. This has led to the field of ANN research. It is recognized that a feed-forward neural network with two or more layers has greater processing ability than a single perceptron. An adaption of the perceptron is the one-layer delta rule introduced by Widrow and Hoff (1960), in which supervised learning is achieved by means of a least mean square (LMS) algorithm. Minsky and Papert (1969) used a simplified perceptron model to investigate the computational capabilities of weighted networks. Early experiments with Rosenblatt's model had aroused unrealistic expectations in some quarters, and there was no clear understanding of the class of pattern recognition problem which it could solve efficiently.

In particular, Minsky and Papert (1969) analyzed the features and limitations of the perceptron model rigorously. The question had arisen of whether a given set of nonlinearly separable patterns can be decomposed in such a way that the largest linearly separable subset can be detected. In particular, Minsky and Papert indicated that a perceptron could not learn functions that are not linearly separable. Neural network research declined throughout the 1970s until the mid-1980s because of an inability to find efficient methods to solve such nonlinearly separable problems. In the 1980s, there was fresh motivation in neural networks research as a result of the increase in computing power and the development of several new algorithms. Hopfield (1982) applied a particular nonlinear dynamic structure to solve problems in optimization. The back-propagation algorithm developed by Rumelhart et al. (1986) gave a strong impulse to the subsequent research, and resulted in the largest body of research and applications in ANNs.

9.3 Basic Architecture of a Multi-layer Perceptron

MLPs, known as feed-forward neural networks, are the most popular and most widely used models in many practical applications. The feed-forward concept indicates that an input pattern is presented to the network via the input layer, and the input signals are passed to the nodes in the next layer in a feed-forward manner. The network is divided into layers. The input layer consists of the inputs to the network, followed by a hidden layer consisting of neurons or hidden units, and then an output layer representing classes or patterns. Each neuron performs a weighted summation of the inputs, which then passes a nonlinear activation function. This summation of the output is called the output layer. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer. Graphically, a neural network can be thought of as a diagram that illustrates how computations are implemented (Fig. 9.1). The diagram includes circles called neurons and lines called connections.

MLPs have the ability to learn through a training process that is described as "supervised," i.e., they must be taught the characteristics of the data set of a particular study. Training requires a set of training data consisting of a series of input and output variables. The objective of training is to find the combination of appropriate connection weights that results in the smallest error of the MLP. During training, the MLP is repeatedly presented with the training data, and the weights in the network are adjusted until the desired input–output mapping occurs. Once trained with representative training data, the MLP can generalize unseen input data.

Training is critical in order to learn the nature of the particular data set under study. The training algorithm is the most important component of an MLP application. A number of algorithms were introduced to train MLP networks, such as the Levenberg–Marquardt technique (Seber and Wild 1989) and back-fitting (Ghosh and Bose 2004). The Levenberg–Marquardt method improves the rate of convergence, but it requires a huge memory space. The back-propagation algorithm, which

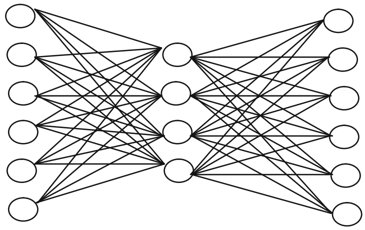


Fig. 9.1 A simple three-layer MLP neural network

is known as the generalized delta rule or gradient descent training rule, is the most popular. The working of the algorithm can be explained by forward and backward steps. In the forward step, data from the input layer are presented to the hidden layer and propagated forward to estimate the output value for each training pattern set. As the forward step is completed, the output is compared with the desired values, and the error is computed. The error is then propagated backward through the network, and the weights are corrected according to the generalized delta rule.

During the training process, the computations in the layers of the neural networks are implemented from the hidden to the output layers. They are not implemented in the input layer. The computation in a given hidden layer can be expressed as

$$\text{net}_j = \sum w_{ij} x_i \tag{9.1}$$

where net_j is the weighted sum of a hidden layer node, w_{ij} represents the connection weight from neuron i (input variable) to neuron j (hidden layer), and x_i is the input variable. This weighted sum is known as activation in the field of ANNs. It is then transformed using a differentiable continuous nonlinear activation function to give the output of the hidden nodes. Different activation functions can be chosen for different types of problems. The sigmoidal function, mostly used in nonlinear transfer functions, is used in the computation in hidden neurons before release to the output layer neuron. Other activation functions which can be employed are logistic functions, threshold functions, a tanh function, and a simple linear activation (identity function). The calculation can be expressed as

$$o_j = \frac{1}{1 + e^{-\text{net}_i}} \tag{9.2}$$

The same procedure is applied to the next layer, where the hidden layer output values are summed to produce the final results in the output layer.

The connection weights are updated during the training process according to the generalized delta rule (Rumelhart et al. 1986). For example, (9.3) is the equation for updating connection weights from the input layer to the hidden layer.

$$\Delta w_{ji(n+1)} = \eta(\delta_j o_i) + \alpha \Delta w_{ji(n)} \tag{9.3}$$

where $\Delta w_{ji(n+1)}$ is the change of a weight from neuron i to neuron j at the $(n+1)$-th iteration, η is the learning rate, δ_j is an index of the rate of change of the error with respect to output from neuron j, and α is a momentum term. In the same way, connection weights between the hidden and output layer are also updated to minimize the errors of the network outputs.

It should be noted that an MLP does not provide a direct means of solving a problem. An MLP needs to be trained with sample data sets before it can be usefully applied. The training of networks is the process of teaching the networks to distinguish input patterns and output responses. It can be described as supervised because this implies we have prior knowledge about the nature of the solution and are supervising the training. Calibration data may be divided into two types of training and testing samples. Training samples are employed to train the network. On the other hand, testing samples, which the network has not previously seen, are used to evaluate the performance of the network. Several factors influence the capacities of the network to generalize and interpolate data that the network has not previously seen. In particular, the key factors are the number of nodes and architecture, the size of training samples, and the learning rate.

A back-propagation algorithm is most commonly used to train MLPs, but it is often too slow for practical applications. The drawbacks of this algorithm have led to a large amount of research into finding fast training algorithms, especially heuristic methods via an adjustable learning rate and a momentum parameter (Yam et al. 1997). An improvement in the convergence rate of the back-propagation algorithm has been sought by several researchers. The learning rate and momentum can be adjusted dynamically by a fixed step based on error observation (Jacobs 1988). Ooyen and Neinhuis (1992) proposed a different error cost function and the use of a second-order Newton's method to optimize these terms. A dynamic learning rate and momentum optimization of the back-propagation algorithm using derivatives with respect to the learning parameters have been considered (Yu and Chen 1997; Yu et al. 1995).

Many approaches to investigating learning rate and momentum have been presented. A genetic algorithm for self-adaptation to speed up the steepest descent rate has been introduced. The key idea is to increase and decrease the learning rate slightly to evaluate the cost function for different values of the learning rate, and then choose the one that has the lowest value of the cost function (Salomon and Hemmen 1996). Karras and Perantonis (1995) presented a Lagrange multiplier approach to the minimization of the cost function in order to improve convergence. Yam and Chow (2000) proposed an approach to finding the optimal weights of a feed-forward neural network. The rationale behind this approach was to reduce the initial network error whilst preventing the network from getting stuck with initial weights. Kamarthi and Pittner (1999) proposed an algorithm based on the extrapolation of each individual interconnection weight to accelerate the back-propagation (BP) algorithm. This requires the error surface to have a smooth variation along the respective axes, so that extrapolation is possible. To perform extrapolation, the BP algorithm convergence behavior of each network weight is examined individually at the end of each iteration. Cho and Chow (1999) presented an approach based on the least-squares method to determine the weights between the output layer and the hidden layer in order to maintain convergence. During problems of local minima, a penalty function optimization method was employed.

9.4 MLP Applications in Land Cover Classification

Parametric statistical methods, i.e., maximum likelihood and Mahalanobis distance, employed for supervised land cover classification depend on an assumption of the Gaussian distribution of the input data. Each class in the feature space is assumed to have an *n*-dimensional multi-variate Gaussian distribution. In reality, input data may not follow the assumed model. Interest in the use of MLPs is a result of their freedom from assumptions about the form and distribution of the input data, their ability to generate nonlinear decision boundaries, and their ability to generalize inputs as well as to learn complex patterns. The statistical methods depend on the assumed model, but a neural network depends on data. In addition, neural networks are suitable for integrating data from different sources.

MLP networks, which are a supervised classification method, have been applied for a variety of applications in diverse fields. All applications of MLPs can be categorized as pattern classification, prediction, and function approximation. One of the most important applications of an MLP is pattern classification. Pattern classification involves classifying data into discrete classes. This section focuses on the classification of land use/cover based on MLPs. The MLPs have a significant role for land use/cover classification because they can handle massive, complex, and incomplete data sets efficiently and produce results showing the accuracy of the classification. MLPs have been applied to the classification of remotely sensed data to distinguish land cover types (Key et al. 1989; Kanellopoulos et al. 1990; Benediktsson et al. 1990; Hepner et al. 1990; Civco 1993; Dreyer 1993; Paola and Schowengerdt 1995). These applications have indicated that neural networks have the adaptability and capacity to produce classifications with a higher level of accuracy than conventional statistical methods.

MLPs are a supervised classification method, and therefore the training algorithm is the key to MLP application. The ultimate goal of training is to generalize outside the training set and predict outputs from unseen input pixels. From a survey of previous studies, it is apparent that a three-layer MLP can be sufficient for land cover classification (Kurkova 1992). The first task when applying an MLP is the identification of the structure of a network, which refers to the topological arrangement of the nodes. The structure of a neural network is determined by the number of input and output nodes. It is particularly important to generalize, interpolate, and extrapolate the data under study, as this affects the capacities of the neural network. Specifically, the application of an MLP involves the setup of the network topology, including the sizes of the input layer, the hidden layer, and the output layer. The number of input layer nodes in an MLP represents input variables, i.e., spectral bands of satellite imagery and ancillary variables. The determination of the variables presented to the input layer is critical. A large number of inputs may reduce the network's generalization and cause redundant information, while a small number of input variables could be insufficient for the network to learn the nature of the training data (Kavzoglu and Mather 2002). The number of output layer nodes determines the complexity of the neural network model. In principle, the greater the number of output nodes to be delineated, the more difficult the problem may be because of the separation of the input space into more specific patterns.

The number of hidden layers and nodes in the hidden layer determines the complexity and power of an ANN to delineate patterns inherent in a particular data set. Kurkova (1992) indicated that a single hidden layer is sufficient for most problems, including classification tasks. An MLP model with a single hidden layer having sigmoidal transformation can approximate any continuous function to any given level of accuracy if a sufficiently large number of hidden nodes are used (Hornik et al. 1989). However, some researchers have reported that the use of two hidden layers in the network for land cover classification can be beneficial (Berberoglu et al. 2007; Aitkenhead and Aalders 2008). The number of hidden nodes affects both classification accuracy and the time of training. A small number of hidden nodes may not identify sufficient internal patterns in the data, and therefore may produce lower classification accuracy. On the other hand, too large a number may be over-specific to the training data. This is known as over-fitting. Identifying the optimum number of hidden layer nodes is not easy because it involves input and output units, the number of training samples, the complexity of the classification to be learned, the level of noise in the data, the network architecture, the nature of the hidden unit activation function, the training algorithm, and regularization (Sarle 2000).

Different rules can be used to determine the optimum number of hidden layer nodes. Most rules are based on a function of the numbers of input and output layer nodes, but none of them has been universally accepted. Hecht-Nielsen (1987) states that any continuous function of n variables can be represented by the superposition of a set of $2n+1$ univariate functions, to suggest that any function can be implemented in a single hidden layer neural network having $2N_{\mathrm{i}}+1$ nodes in the single hidden layer, where N_{i} represents the number of input nodes. Paola (1994) derived a formula by making the number of parameters necessary for neural networks equal to the number of parameters required by the maximum likelihood classifier. The formula is given as

$$\frac{2+N_{\mathrm{o}}N_{\mathrm{i}}+\frac{1}{2}N_{\mathrm{o}}\left(N_{\mathrm{i}}^{2}-N_{\mathrm{i}}\right)-3}{N_{\mathrm{i}}+N_{\mathrm{o}}} \tag{9.4}$$

where N_{i} is the number of input nodes, and N_{o} is the number of output nodes. Other authors proposed different heuristic formulas. For example, Hush (1989) proposed a heuristic rule of $3N_{\mathrm{i}}$, Ripley (1993) proposed the formula $N_{\mathrm{i}}+N_{\mathrm{o}}/2$, Wang (1994) suggested the formula $2N_{\mathrm{i}}/3$, Kanellopoulos and Wilkinson (1997) used $2N_{\mathrm{i}}$ or $3N_{\mathrm{i}}$, and Garson (1998) employed $N_{\mathrm{p}}/[r(N_{\mathrm{i}}+N_{\mathrm{o}})]$ for computing the number of hidden layer nodes, where N_p is the number of training samples, and r is a constant relating to the noise level of the data, which ranges from 5 to 10.

The other important training parameters that need to be examined are initial weight values, learning rate and momentum, and the size of training samples. There is no genetic formula that can be used to choose such parameter values, so the parameters are usually determined by trials. These parameters are adjusted and modified during the training of the network.

The initial weights of the network play a significant role in the convergence of the training method. In addition, weight initialization has been known as one of the most effective approaches in accelerating the training of a neural network (Drago and Ridella 1992; Martens 1996). Without a priori information about the final weights, it is common practice to initialize all weights randomly with a small absolute value. If large initial values are assigned to the weights, the neurons in the network are driven to saturation. In this case, local gradients on the error surface assume small values, causing the learning process to slow down. On the other hand, if small initial weight values are assigned, the algorithm of back-propagation may operate on a very flat area around the origin of the error surface. Hence, the use of both large and small initial weight values should be avoided. There is no universally accepted method for determining an optimum range for initial weight values.

The learning rate and the momentum factor control the size of the weight adjustment in the descent direction and the dampening of the oscillations of the iterations. Jacobs (1988) presented a simple method for updating learning rate and momentum. He suggested dynamically increasing or decreasing learning rate and momentum by a fixed factor based on observations of the error signals. Variations of Jacob's method were reported by Vogel et al. (1988). Many heuristics of learning rate and momentum have been proposed in previous studies using different data sets (Paola and Schowengerdt 1997; Gong 1996; Staufer and Fisher 1997). In addition, more advanced and sophisticated methods have been developed to identify the optimum values for learning rate and momentum. These methods adapt the learning rate during the training process in relation to the characteristics of the error surface and gradient. Such methods, as in Heermann and Khazenie (1992), are known as adaptive learning strategies.

The size of the training sample plays a vital role in the performance of a supervised neural network. The extraction of land cover information from satellite images involves the selection of training data and classification techniques. The training stage of the classification process is particularly important because it has a significant impact on the performance of any classification technique. This is especially important for MLP networks, as they learn the characteristics of the data from sample values, and identify the pixels in the image remainder. Both the quality and size of the training data are of critical importance for a successful classification (Kavzoglu 2008). A small sample size is not enough for a neural network to recognize all classes and identify the class boundaries in the input space. On the other hand, a large number of training samples make the network over-specific and require more time for training (Kavzoglu 2001). The number of training samples should be defined for each class according to the class complexity (Blamire 1996).

The impacts of training size on the performance of land cover classification have been reported by authors such as Foody et al. (2006) and Kavzoglu (2008). Several attempts have been made to estimate the appropriate size of the training sample with respect to the network topology and the expected accuracy of the result. For a standard statistical classifier, such as a maximum livelihood classifier, 30p pixels for one class should be used, where p is the number of input variables (Mather 1999). This heuristic rule may not be feasible for neural networks because more training samples are needed to learn the characteristics of a class. Therefore, different heuristic rules are proposed to estimate the number of training samples according to a given network topology (the number of input, hidden, and output nodes) and the connection weights of the network. Hush (1989) proposed that a minimum training size of $30N_i(N_i+1)$ should be employed, where N_i is the number of input variables. Garson (1998) introduced three heuristic rules. The first is that the number of training samples should be at least ten times the number of input variables. The second stated that the number of training samples should be at least ten times the number of nodes in the input and hidden layers of the network. The third indicated that 30 times as many input patterns as network weights should be employed to avoid over-fitting. However, Baum and Haussler (1989) proposed that $10N_w$ (where N_w is the totality of the number of weights in a given network) should be employed to estimate the number of training samples. Klimasauskas (1993) reported that a training size of $5N_w$ should be used.

9.5 Future Directions

MLP networks have been a useful approach because they have their own strengths compared with conventional methods such as statistical regression, pattern recognition, and time series analysis. Therefore, they can be employed as a powerful alternative to such methods. However, they also have disadvantages. MLPs are like a black box in nature. If the problem is to find output patterns to input, MLP models are an appropriate tool. In cases where a casual relationship between input and output needs to be determined, statistical methods may be more suitable than MLPs. Another disadvantage of MLPs is that the training process consumes too much time because it requires determining the network structure and adjusting the connection weights. In particular, MLP training takes a lot of computer memory, and it may take several hours before the network converges to a minimum error point. Statistical regression methods may produce results in a shorter time.

Although drawbacks exist in the use of MLPs, this approach will play an important role in various applications, including the geospatial analysis domain, because of the advancement of computer processing power. The theoretical developments and advances in MLPs are the result of research efforts in the development of new training algorithms and the combined use of the MLP with other computing methods such as fuzzy logic, genetic algorithms, and decision trees. The combination may allow an MLP system to take advantage of both of the paradigms and overcome the limitations of MLPs.

MLPs have been widely used in many areas. The most commonly used method to train the MLP is based on the back-propagation algorithm. Many variations of this standard algorithm have been proposed and new ones continue to come out regularly. Some of the popular methods are the Delta-Bar-Bar, Vogl, Rprop, SuperSAB, Quickprop, and Levenberg–Marquardt algorithms (Hagan and Menhaj 1994; Bishop 1995). New algorithms have been proposed to solve issues regarding the slow convergence of the back-propagation training algorithm and local minima entrapment.

Several methods have been proposed to speed up the back-propagation-based training algorithms by fixing an appropriate learning rate and momentum value for each layer at the time of training (Yam and Chow 1997). Different initialization techniques (Yam and Chow 2000; Yam and Chow 2001) and cost optimization techniques (Kwok and Yeung 1997) have been proposed to increase the rate of convergence. Dynamic tunneling (Chowdhury et al. 1999) is a technique used to de-trap local minima. Abid et al. (2001) have described a modified standard back-propagation algorithm using the sum of the squares of the linear and nonlinear errors for all output units and for the current pattern. Yam and Chow (1997) have proposed an extended least-squares-based algorithm for training feed-forward networks. In this, the weights connecting the last hidden and the output layers are first evaluated by a least-squares algorithm. The weights between the input and the hidden layers are then evaluated using a modified gradient descent algorithm. Kwok and Yeung (1997) have studied different objective functions for training new hidden units in constructive neural networks. They have followed a layer-by-layer optimization technique.

It is critical to decide a learning rate to speed up the convergence of the global minimum of the mean square of the network outputs. A small learning rate results in a slow training progress. On the other hand, a large learning rate accelerates the progress significantly. Several studies have proven that the use of a momentum term in the algorithm can be useful in accelerating the convergence and avoiding local minima. The momentum is defined as a fraction of the previous weight change. The momentum is employed to stabilize weight change using a combination of a decreasing gradient term with a fraction of the previous weight change. An improvement in the training algorithm by the use of the momentum is a further development. For example, Yu et al. (1993) developed an adaptive momentum algorithm that updates the momentum at each iteration. The results of simulations have indicated that this method can remove possible divergent oscillations during initial training, speed up the learning process, and produce a lower error at the final convergence stage. Jeenbekov and Sarybaeva (2000) characterized the properties of various parameters of the sigmoidal function, and they analyzed the effect on the speed of convergence in training an MLP with the back-propagation algorithm. Wang et al. (2004) proposed an improved back-propagation algorithm to avoid local minima caused by neuron saturation in the hidden layer. If the network does not produce the desired results, an activation function is adapted to prevent saturation in the hidden layer. Bi et al. (2005) proposed a modified error function with two terms. By adding one term to the conventional error function, a modified error function can harmonize the update of weights connected to the hidden layer as well as those connected to the output layer. Thus, it can avoid the local minima problem caused by update disharmony between weights connected to the hidden layer and the output layer. Simulations on some benchmark problems and a real classification task have been performed to test the validity of the modified error function.

Another aspect of the advances is the combination of fuzzy logic with the MLP networks. The power of neural networks is combined with fuzzy logic to enable fuzzy rules to be incorporated into the pattern classification. Fuzzy logic is capable of interpreting imprecise data by making decisions possible via linguistic rules. The MLP networks are able to learn with a sufficient number of observed samples. The capabilities and restrictions of each method led to the development of a hybrid fuzzy–neural method. A number of different schemes and architectures of the hybrid method have been proposed, such as fuzzy neurons (Gupta 1994), neural networks with fuzzy weights (Buckley and Hayashi 1994), neuro–fuzzy adaptive models (Brown and Harris 1994), and fuzzy-logic-based neurons (Pedrycz 1995). Takagi and Hayashi (1991) proposed a neural–fuzzy reasoning system that is capable of automatic determination of inference rules and adjustments according to the time-variant reasoning environment with the use of a neural network in fuzzy reasoning. Horikawa et al. (1992) presented a fuzzy modeling method using fuzzy neural networks with the BP algorithm. This method can automatically identify the fuzzy model of a nonlinear system. Nie and Linkens (1992) approximated reasoning through a back-propagation neural network with the aid of fuzzy set theory.

References

Abid S, Fnaiech F, Najim M (2001) A fast feedforward training algorithm using a modified form of the standard backpropagation algorithm. IEEE Trans Neural Network 12:424–430

Aitkenhead MJ, Aalders IH (2008) Classification of Landsat Thematic Mapper imagery for land cover using neural networks. Int J Rem Sens 29:2075–2084

Atkinson PM, Tatnall ARL (1997) Neural networks in remote sensing. Int J Rem Sens 18:699–709

Baum EB, Haussler D (1989) What size net gives valid generalization? In: Touretzky DS (ed) Advances in neural information processing systems I. Morgan Kaufmann, San Mateo, pp 81–90

Benediktsson JA, Swain PH, Ersoy OK (1990) Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Trans Geosci Rem Sens 28:540–552

Berberoglu S, Curran PJ, Lloyd CD, Atkinson PM (2007) Texture classification of Mediterranean land cover. Int J Appl Earth Observation Geoinformation 9:322–334

Bi W, Wang X, Tang Z, Tamura H (2005) Avoiding the local minima problem in backpropagation algorithm with modified error function. Trans Fund Electron Comm Comput Sci E88-A:3645–3653

Bishop C (1995) Neural networks for pattern recognition. Clarendon, Oxford

Blamire PA (1996) The influence of relative sample size in training artificial neural networks. Int J Rem Sens 17:223–230

Brown M, Harris C (1994) Neuro-fuzzy adaptive modelling and control. PrenticeHall, New York

Buckley JJ, Hayashi Y (1994) Fuzzy neural networks. In: Yager R, Zadeh L (eds) Fuzzy sets, neural networks and soft computing. Van Nostrand Reinhold, New York

Cho SW, Chow TWS (1999) Training multilayer neural networks using fast global learning algorithm-least-squares and penalized optimization methods. Neurocomputing 25:115–131

Chowdhury PR, Singh YP, Chansarkar RA (1999) Dynamic tunneling technique for efficient training of multilayer perceptrons. IEEE Trans Neural Network 10:48–55

Civco DL (1993) Artificial neural networks for land cover classification and mapping. Int J Geogr Inform Syst 7:173–186

Cronin J (1987) Mathematical aspects of Hodgkin–Huxley theory. Cambridge University Press, Cambridge

Drago GP, Ridella S (1992) Statistically controlled activation weight initialization. IEEE Trans Neural Network 3:627–631

Dreyer P (1993) Classification of land cover using optimized neural networks on SPOT data. Photogramm Eng Rem Sens 5:617–621

Foody GM, Mathur A, Sanchez-Hernandez C, Boyd DS (2006) Training set size requirements for the classification of a specific class. Rem Sens Environ 104:1–14

Garson GD (1998) Neural networks: an introductory guide for social scientists. Sage, London

Ghosh AK, Bose S (2004) Backfitting neural networks. Comput Stat 19:193–210

Girosi F, Poggio T (1990) Networks and the best approximation property. Biol Cybern 63:169–176

Gong P (1996) Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping. Photogramm Eng Rem Sens 62:513–523

Gupta MM (1994) Fuzzy neural networks: theory and Applications, Proceedings of SPIE, pp 303–325

Hagan MT, Menhaj MB (1994) Training feedforward neural networks with the Marquardt algorithm. IEEE Trans Neural Network 5:989–993

Hartigan JA, Wong MA (1979) A k-means clustering algorithm. Appl Stat 28:100–108

Hebb DO (1949) The organization of behavior. Wiley, New York

Hecht-Nielsen R (1987) Kolmogorov's mapping neural network existence theorem. In: Caudill M, Butler C (eds) Proceedings of the first IEEE international conference on neural networks, pp 11–14

Heermann PD, Khazenie N (1992) Classification of multispectral remote sensing data using a back-propagation neural network, LEEE Transactions on Geoscience and Remote Sensing, 30(1):81–88

Hepner GF, Logan T, Ritter N, Bryant N (1990) ANN classification using a minimal training set: comparison to conventional supervised classification. Photogramm Eng Rem Sens 56:469–473

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79(8):2554–2558

Horikawa S, Furuhashi T, Uchikawa Y (1992) On fuzzy modelling using fuzzy neural networks with backpropagation algorithm. IEEE Trans Neural Network 3:801–806

Hornik K, Stinchcombe M, White H (1989) Multi-layer feedforward networks are universal approximators. Neural Network 2:359–366

Hush DR (1989) Classification with neural networks: a performance analysis. In: Proceedings of the IEEE international conference on systems engineering, pp 277–280

Jacobs RA (1988) Increasing rate of convergence through learning rate adaptation. Neural Network 1:295–307

Jeenbekov AA, Sarybaeva AA (2000) Conditions of convergence of backpropagation learning algorithm. In: Proceedings of SPIE on optoelectronic and hybrid optical/digital systems for image and signal processing, pp 12–18

Kamarthi SVC, Pittner S (1999) Accelerating neural network training using weight extrapolations. Neural Network 12:1285–1299

Kanellopoulos I, Wilkinson GG (1997) Strategies and best practice for neural network image classification. Int J Rem Sens 18:711–725

Kanellopoulos I, Varfis A, Wilkinson GG, Megier J (1990) Land-cover discrimination in SPOT HRV imagery using an artificial neural network a 20-class experiment. Int J Rem Sens 13:917–924

Karras DA, Perantonis SJ (1995) An Efficient constrained training algorithm for feedforward networks. IEEE Transactions on Neural Networks 6:1420–1434

Kavzoglu T (2001) An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images. Ph.D. Dissertation, The University of Nottingham

Kavzoglu T (2008) Increasing the accuracy of neural network classification using refined training data. Environ Model Softw 24:850–858

Kavzoglu T, Mather PM (2002) The role of feature selection in artificial neural network applications. International Journal of Remote Sensing 23:2919–2937

Key J, Maslanik JA, Schweiger AJ (1989) Classification of merged AVHRR and SMMR Arctic data with neural networks. Photogramm Eng Rem Sens 55:1331–1338

Kim DS, Nelson RF (1998) Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. Int J Rem Sens 19:2639–2663

Klimasauskas CC (1993) Applying neural networks. In: Trippi RR, Turban E (eds) Neural networks in finance and investing. Probus, Cambridge, pp 47–72

Kohonen T (1984) Self-organization and associative memory. Springer, Berlin

Kohonen T (1995) Self-organizing maps. Spinger, Berlin

Kohonen T (1990) The self-organizing map. Proc. IEEE 78:1464–1480

Kurkova V (1992) Kolmogorov's theorem and multilayer neural networks. Neural Network 5:501–506

Kwok TY, Yeung DY (1997) Objective functions for training new hidden units in constructive neural networks. IEEE Trans Neural Network 8:1131–1147

Martens JP (1996) A stochastically motivated random initialization of pattern classifying MLPs. Neural Process Lett 3:23–29

Mather PM (1999) Computer processing of remotely-sensed images: an introduction. John Wiley, Chichester

McCulloch WW, Pitts W (1943) A logical calculus of ideas imminent in nervous activity. Bull Math Biophys 5:115–133

Mead C (1989) Analog VLSI and neural systems. Addison-Wesley, Reading

Minsky M, Papert SA (1969) Perceptrons: an introduction to computational geometry. MIT Press, Cambridge

Nie J, Linkens D (1992) Neural network-based approximate reasoning: principles and implementation. Int J Control 56:399–413

Ooyen AO, Neinhuis B (1992) Improving the convergence of the backpropagation algorithm. Neural Network 5:465–471

Paola JD (1994) Neural network classification of multispectral imagery. M.Sc. Thesis, The University of Arizona

Paola JD, Schowengerdt RA (1995) A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification. IEEE Trans Geosci Rem Sens 33:981–996

Paola JD, Schowengerdt RA (1997) The effect of neural network structure on a multispectral land-use/land-cover classification. Photogramm Eng Rem Sens 63:535–544

Pedrycz W (1995) Fuzzy sets engineering. CRC Press, Boca Raton

Poggio T, Girosi F (1987) Networks for approximation and learning. In: Proceedings of IEEE, pp 1481–1497

Ripley BD (1993) Statistical aspects of neural networks. In: Barndorff-Nielsen OE, Jensen JL, Kendall WS (eds) Networks and chaos: statistical and probabilistic aspect. Chapman & Hall, London, pp 40–123

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructures of cognition. MIT Press, Cambridge, pp 318–362

Salomon R, Hemmen JL (1996) Accelerating backpropagation through dynamic self-adaptation. Neural Network 9:589–601

Sarle WS (2000) Neural network, FAQ. ftp://ftp.sas.com/pub/neural/FAQ.html

Seber GAF, Wild CJ (1989) Nonlinear regression. Wiley, New York

Staufer P, Fisher MM (1997) Spectral pattern recognition by a two-layer perceptron: effects of training set size. In: Kanellopoulos I, Wilkinson GG, Roli F, Austin J (eds) Neurocomputation in remote sensing data analysis. Springer, London, pp 105–116

Takagi T, Hayashi I (1991) Neural network driven fuzzy reasoning. Int J Approx Reason 5:191–212

Vogel TP, Mangis JK, Rigler AK, Zink WT, Alkon DL (1988) Accelerating the convergence of the backpropagation method. Biol Cybern 59:257–263

Wang J (1994) A deterministic annealing neural network for Conex Programming. Neural networks 7:629–641

Wang XG, Tang Z, Tamura H, Ishii M, Sun WD (2004) An improved backpropagation algorithm to avoid the local minima problem. Neurocomputing 56:455–460

Widrow B, Hoff ME (1960) Adaptive switching circuits. IRE WESCOM Convention Record 4:96–104

Yam JYF, Chow TWS (1997) Extended least squares based algorithm for training feedforward networks. IEEE Trans Neural Network 8:806–810

Yam JYF, Chow TWS (2000) A weight initialization method for improving training speed in feedforward neural networks. Neurocomputing 30:219–232

Yam JYF, Chow TWS (2001) Feedforward networks training speed enhancement by optimal initialization of the synaptic coefficients. IEEE Trans Neural Network 12:430–434

Yam JYF, Chow TWS, Leung CT (1997) A new method in determining initial weights of feedforward neural networks for training enhancement. Neurocomputing 16:23–32

Yu XH, Chen GA (1997) Efficient backpropagation learning using optimal learning rate and momentum. Neural Network 10:517–527

Yu X, Loh NK, Miller WC (1993) A new acceleration technique for the backpropagation algorithm. In: Proceedings of IEEE international conference on neural networks, pp 1157–1161

Yu XH, Chen GA, Cheng DSX (1995) Dynamic learning rate optimization of the backpropagation algorithm. IEEE Trans Neural Network 6:669–677

Zhu Y, He Y (2006) Short-term load forecasting model using fuzzy c means based radial basis function network. In: Proceedings of 6th international conference on intelligence systems design and applications, pp 579–582

Chapter 10
Agent-Based Simulation in Geospatial Analysis

Kondwani Godwin Munthali

10.1 Introduction

There is a wide array of simulation methods that mimic the mechanisms of human intelligence to achieve one or more objectives. Analytical simulation approaches basically use equations that explain data, while statistical ones work primarily with probabilities. An iterative combination of any or both of the above uses feedback options to answer problems which are too complex to be solved by one equation. Most of these equation-based mathematical models identify system variables, and evaluate or integrate sets of equations relating to these variables. A variant of such equation-based models are based on linear programming (Howitt 1995; Weinberg et al. 1993), and are potentially linked to geographical information science (GIS) information (Chuvieco 1993; Cromley and Hanink 1999; Longley et al. 1994). However, in practice there are limited levels of complexity that can be built into these models (Parker et al. 2003).

To incorporate complexity, sets of differential equations linked through intermediary functions and data structures are sometimes used to represent stocks and flows of information (Gilbert and Troitzsch 1999). Although they include human and ecological interactions, these systemic models tend to have difficulties in accommodating spatial relationships (Baker 1989; Sklar and Costanza 1991). Given their power and ease of use, statistical simulation approaches have been widely accepted, largely because they include a variety of regression techniques applied to space and more tailored spatial statistical methods (Ludeke et al. 1990; Mertens and Lambin 1997). However, according to Parker et al. (2003), unless tied to theoretical frameworks, statistical models tend to down-play decision-making and social phenomena. Other simulation approaches express qualitative knowledge in a quantitative fashion by

K.G. Munthali (✉)
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kmunthali@gmail.com

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_10, © Springer Japan 2012

combining expert judgement with probability techniques such as Bayesian or artificial intelligence approaches (Parker et al. 2003).

The gaps and inconsistencies left by these modeling approaches saw the proliferation of cellular automata (CA) in combination with Markov models. In CA, each cell exists in one of a finite set of states, and future states depend on transition rules based on a local spatio-temporal neighborhood (Kamusoko et al. 2009), while in Markov models, cell states depend probabilistically on temporally lagged cell state values. These cellular models (CMs) underlie many land-use studies in which Markov–CA combinations are common (Balzter et al. 1998; Li and Reynolds 1997; Kamusoko et al. 2009). While many CMs assume that the actions of human agents are important, and others assume a set of agents coincident with lattice cells and use transition rules as proxies to decision-making, they both fail to simulate decisions expressly and explicitly (Parker et al. 2003). In the latter case, the actor is not tied to locations and, as Hogeweg (1988) observed, this introduces problems of spatial orientation to the extent that the intrinsic neighborliness of CA relationships do not reflect on the actual spatial relationships. This highlights the main challenge faced by CMs and most of the aforementioned modeling approaches when it comes to incorporating individualistic human decision-making (Parker et al. 2003). When the focus is on human actions, agents become the crucial components in the model. While cellular models are focused on landscapes and transitions, agent-based models (ABMs) primarily focus on humans and their actions. Therefore, it is not surprising to realize that an ABM is more of a mindset that builds on describing a system from the perspective of its constituent units than a technology.

The benefits of ABMs over other modeling techniques can be expressed in three statements: (1) they capture emergent phenomena; (2) they provide a natural description of a system; (3) they are flexible. It is clear, however, that the ability of ABMs to deal with emergent phenomena is what drives the other benefits (Bonabeau 2002). Emergent phenomena result from the interactions of individual entities which cannot be reduced to the system's parts: the whole is more than the sum of its parts because of the interactions between the parts (Bonabeau 2002). In the geographical context of level and scale, Auyang (1998) understands "emergence" as emergent phenomena at one level that constitute the units of interaction, or drivers of change at a higher level.

There is a wide range of literature discussing the application of ABMs in a number of global environmental challenges where agents have been used to represent a number of entities, including atoms, biological cells, animals, people, and organizations (Conte et al. 1997; Epstein and Axtell 1996; Janssen and Jager 2000; Liebrand et al. 1988; Weiss 1999). However, in this chapter we seek to add to the current discussion about ABMs in land-use modeling, some of which follow the conceptual framework shown in Fig. 10.1. The rest of the chapter is as follows. We begin by presenting the history of ABMs, followed by the concepts of agent modeling and the tools available for simulations with a bias towards land-use modeling. We later outline the work carried out so far in agent-based land-use modeling, and discuss a selected set of applications. In conclusion, we discuss the advantages and limitations currently facing ABMs, and try to predict their future use.

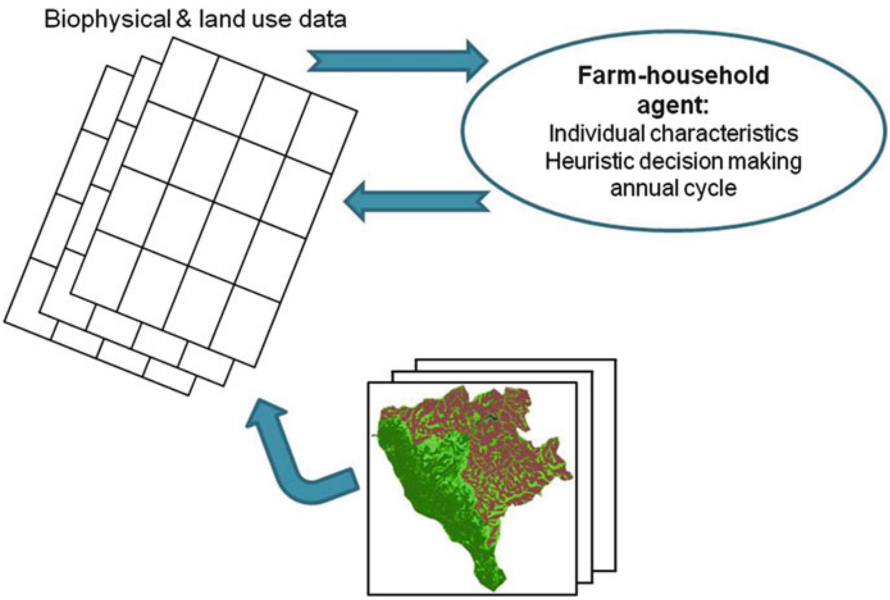


Fig. 10.1 A conceptual framework for a farm-based decision-making ABM (adapted from Deadman et al. 2004)

10.2 History of ABMs

Agent-based modeling can be traced back hundreds of years to discoveries that include Adam Smith's invisible hand in economics, Donald Hebb's cell assembly, and the blind watchmaker in Darwinian evolution (Axelrod and Cohen 2000). In each of these early theories, simple individual entities interact with each other to produce new complex phenomena that seemingly emerge from nowhere (Heath 2010). Because of Newton's reductionist philosophy (Gleick 1987) and his lack of tools to adequately study and understand emergent phenomena, it was not until the theoretical and technological advances were made that led to the invention of the computer that scientists began building models of these complex systems and began to have a better understanding of their behavior (Heath 2010). The pioneering work was carried out by Alan Turing with the invention of the Turing machine around 1937. By replicating any mathematical process, the Turing machine showed that machines were capable of representing real-world systems (Heath 2010). The theoretical scientific belief that machines could recreate the non-linear systems observed in nature got a further boost when Turing and Church later developed the Church–Turing hypothesis, which stated that a machine could duplicate not only the functions of mathematics, but also the functions of nature (Levy 1992). Premised on von Neumann's heuristic use (von Neumann 1966) these machines have since moved from theoretical ideas to the real computers that we are familiar with today (Heath 2010).

Now that computers had come to stay, the scientific focus shifted towards synthesizing the complexity of natural systems. Influenced by a reductionist philosophy, most scientists took a top-down approach (Heath 2010). Evidence of this is seen in early applications of artificial intelligence, where the focus was more on defining the rules of the appearance of intelligence and creating intelligent solutions than focusing on the structure that creates intelligence (Casti 1995). This approach was skewed towards the idea that systems are linear, and thus it failed to enhance our understanding of the complex non-linear systems found in nature (Langton 1989). A U-turn towards a bottom-up approach followed when Ulam suggested that von Neumann's self-reproducing machine could be represented more easily by using cellular automata (CA) (Langton 1989). CA are self-operating entities that exist in individual cells which are adjacent to one another in a 2D space like a checkerboard, and have the capability to interact with the cells around them. According to Heath (2010), the impact of the CA approach was overwhelming for two reasons: (1) because the cells in CA act autonomously and simultaneously with other cells in the system, the simulation process changed from serial to parallel representation, and (2) CA systems are composed of many locally controlled cells that together create global behavior. The former was important because many natural systems are widely accepted to be parallel systems (von Neumann 1966), while the latter led to the bottom-up approach as the CA architecture requires engineering a cell's logic at the local level in the hope that it will create the desired global behavior (Langton 1989).

After learning how to synthesize complex systems and discovering some of their properties using CA, complex adaptive systems (CASs) began to emerge as the direct historical roots of ABMs (Heath 2010). Drawing much of its inspiration from biological systems, CASs were mainly concerned with how complex adaptive behavior emerges in nature from interactions among autonomous agents (Dawid and Dermietzel 2006). Much of the early work in defining and designing CASs resulted from Holland's work to identify properties and mechanisms that compose all ABMs as we know them today (Buchta et al. 2003). Holland reported the three main properties of CASs to be aggregation, non-linearity, which is the idea that the whole system output is greater than the sum of the individual component outputs, and diversity, meaning that agents do not all act the same way when stimulated by a set of conditions.

It is evident that ABMs emerged from the scientific search to try and understand non-linear systems, and this revelation suggests why ABMs are a useful research tool. In summary, many subject areas played an important role in developing the multidisciplinary field of ABMs.

10.3 Agent Modeling

Parker and Meretsky (2004) noted that ABMs often model complex dynamic systems and focus on the macro-scale, or "emergent," phenomena that result from the decentralized decisions of, and interactions between, the agents. The concept behind

ABMs, which was borrowed from the computer sciences, is to mimic human- or animal-like agents interacting at the micro-scale in a computer simulation in order to study how their aggregation leads to complex macro-behavior and phenomena (Berger 2001).

ABMs build on a successful specification of the agent itself, its behavior, the representation of the environment and the interactions. The term agent refers to any individual or group of individuals who exist in a given area and are capable of making decisions for themselves or for the given area. Generally, an agent can represent any level of organization (a herd, a village, an institution, etc.) (Verburg 2006). In land-use modeling, these agents couple a human system making land-use decisions with an environmental system represented by a raster grid (Deadman et al. 2004, see Fig. 10.1).

The specification of the behavior of agents demands a proper description of the actual actions of the agents and the basic elements that cause modifications in their environment and in other agents (Bandini et al. 2009). It also demands the provision of mechanisms for the agents to effectively select the actions to be carried out. The mechanism of an agent refers to the internal structure which is responsible for the selection of actions (Russel and Norvig 1995), while the actions of agents pertain to descriptions of the agents' actions, for instance state transformation, environmental modifications, an agent's perception and responsiveness, and the spatial physical displacement of an agent in the environment. The description of the environment of an agent see Weyns et al (2007), for a detailed definition should, among other factors, primarily define and enforce the rules of behavior of an agent, and maintain the internal dynamics of the system to avoid chaos. At the same time, it should also support an agent's perception and localized actions by embedding and supporting access to objects and parts of the system that are not necessarily modeled as agents (Bandini et al. 2009). Interaction is a key aspect in agent design, both with other agents and/or the environment. Several definitions of interaction have been provided, and most of them focus on the ability of agents to engage with the environment and with other agents in a meaningful problem-solving or goal-oriented scheme to achieve particular objectives according to the coordination, cooperation and competition practices of natural phenomena.

These concepts have been the subject of experiments on many platforms, the choice of which tends to depend largely on the researcher's preference, the computation requirements, and the overall objectives of the study. Most ABM platforms follow the "framework and library" paradigm (Railsback et al. 2006). A framework is a set of standard concepts for designing and describing ABMs, while a library is a set of software implementing the framework and providing simulation tools. Without trying to be exhaustive, we present some of the commonly available agent modeling platforms. The earliest of these platforms include the Swarm (Minar et al. 1996, www.swarm.org), whose libraries were written in Objective-C with later up-dates using Java Swarm in order to allow the use of Swarm's Objective-C library in Java (Railsback et al. 2006). The recursive porous agent simulation toolkit (RePast) (Collier 2000; http://repast.sourceforge.net/) was first developed as a Java implementation of Swarm, but has since evolved into a fully fledged stand-alone Java platform.

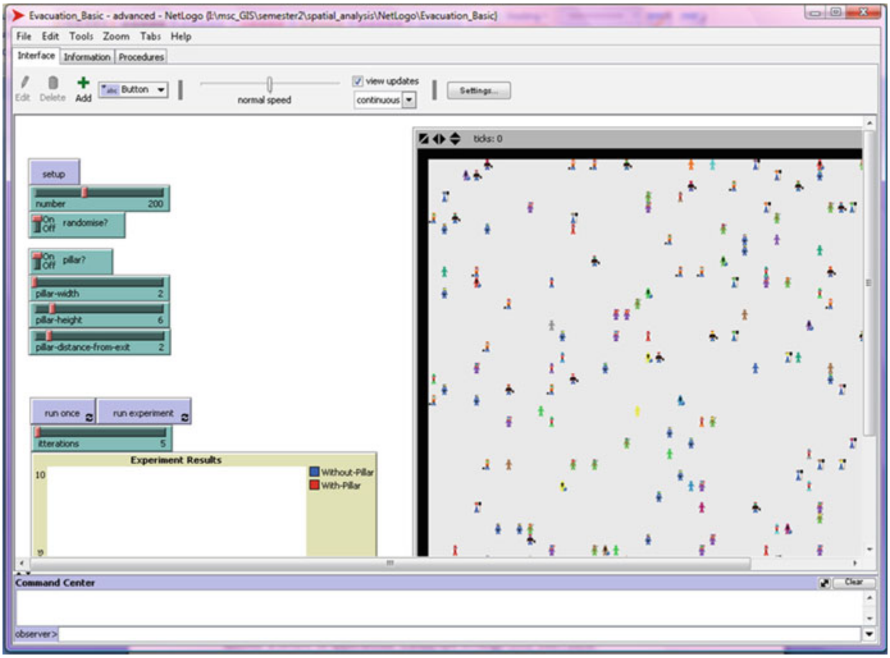


Fig. 10.2 A NetLogo ABM platform

MASON (Luke et al. 2005; http://cs.gmu.edu/~eclab/projects/mason/) was developed later, also as a Java implemented tool. Despite these platforms providing standardized software designs and tools without limiting the type or complexity of the models they implement, they have well-known limitations (Railsback et al. 2006). According to Tobias and Hofmann (2004), their weaknesses include difficulty of use, insufficient tools for building models, and especially tools for representing space, insufficient tools for executing and observing simulation experiments, and a lack of tools for documenting and communicating software. The Logo family evolved from such limitations with the aim of providing a high-level platform that allows model building and learning from simple ABMs (Railsback et al. 2006). Although built on elementary-level principles primarily to aid student learning, NetLogo (http://ccl.northwestern.edu/netlogo/) now contains complex capabilities and is arguably the most widely used platform (Railsback et al. 2006). Figure 10.2 is a screenshot of a NetLogo platform that comes with its own programming language, which is claimed to be simpler to use than Java or Objective-C, an animation display automatically linked to the program, and optional graphical controls and charts.

A model agent is an abstract representation of the real world, the landscape, individuals or groups, and the processes that link these components. Model agents are developed at varying levels of complexity and scales of representation, but their development should offer a level of realism that will not inhibit any validation techniques which will be used later (Deadman et al. 2004). An agent tends to act as an interface in helping to assimilate the broader macro-information into the decision-making process at the grid level, thereby creating an action in response to the natural and economic stimuli (Rajan and Shibasaki 2000). In land-use modeling, the macro-information comes in the form of the biophysical conditions in the area and the prevailing economic conditions at a given location and time.

Mismatches between the units of analysis and the units of actual decision-making have been widely accepted, and attention is slowly shifting from pixels to agents (Verburg et al. 2005). In land use/cover change (LUCC) modeling, for instance, the overarching problem has been linking agents capable of decision-making to land areas: i.e. linking "people and pixels" (Geoghegan et al. 1998; Rindfuss et al. 2003). An expanding group of models has recently used individual agents as units of simulation (see Berger 2001; Bousquet and Le Page 2004; O'Sullivan and Haklay 2000; Parker et al. 2003). While agent-based approaches have specific strengths in describing and exploring decision-making by agents in a variety of fields (see Malleson 2010), they face difficulties in adequately representing the spatial patterns in LUCC models owing to difficulties in representing the feedback between the behavior of the agents and land units (Verburg 2006). In some ABMs, a cellular automata (CA) approach is used, in which the state of a pixel is determined by the state of the neighboring pixels (Ligtenberg et al. 2004; Manson 2005). Although CA methods are often seen as a type of multiagent approach, because of the explicit treatment of interactions between (spatial) entities it is hard to imagine that the pixels are a representation of the agents (Couclelis 2001).

In current practice, a cellular component that represents the landscape is coupled with an agent-based component that represents the human decision making (Schreinemachers and Berger 2006; Parker et al. 2003). As the debate progressively leans towards agents and away from pixels, challenges about how to represent real-world decision making become more apparent. The decision-making structure of an agent falls into two broad categories, optimizing and heuristic. The key difference is that the latter have neither the information to compare all feasible alternatives nor the computational power to select the optimum (Schreinemachers and Berger 2006). Heuristics are relatively simple rules that build on the concept of a search process guided by rational principles (Simon 1957), while optimization needs the ability to process large amounts of information about all feasible alternatives and always select the best one (Schreinemachers and Berger 2006). The intuitive nature of heuristics makes them more transparent and therefore easy to validate. However, constructing a decision tree which is representative of the thought processes of a human being is not easy. A variety of optimization approaches are available, but the most common include mathematical programming (see Balmann 1997; Berger 2001; Becu et al. 2003; Happe 2004) and genetic programming (see Manson 2005). Mathematical programming (MP) is a computerized search for a combination of decisions that yields the highest objective function value (Schreinemachers and Berger 2006). Unlike the heuristic approach, MP requires the explicit specification of an objective function. In LUCC modeling, the objectives of the agents, which include cash income, food, and leisure time, tend to be similar for both MP and heuristic approaches. Figure 10.3 gives an example of a heuristic decision-making tree.

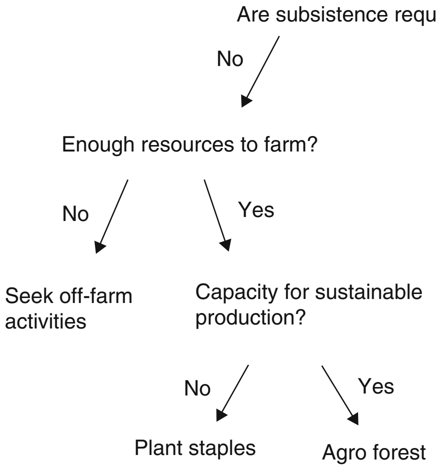


Fig. 10.3 A heuristic structure of subsistence farm-based decision-making (adapted from Deadman et al. 2004)

10.4 ABM Applications

Using a mountainous region in Laos, Wada et al. (2007) developed a micro-scale ABM to simulate the spatial and temporal patterns of shifting cultivation with the aim of understanding how this expands in space. While ABMs recognize and take advantage of the fact that human decision-making is heterogeneous, decentralized and autonomous (Parker et al. 2003), this is a representative case in which individual behavior is conspicuously less heterogeneous and less decentralized. The base unit in the model was a cluster of villages as opposed to individual households (see Deadman et al. 2004; Evans and Kelley 2004). The choice of a cluster of villages in the Laotian model was partly because of the limited availability of spatial data (village boundary data), and also because of the revelation that decisions to expand and/or relocate shifting cultivation are made at village level rather than in individual households (Wada et al. 2007).

Underscoring the sporadic, incomplete and mostly non-existent market context in the subsistence agriculture set-up, Walker (1999) attempted to account for land allocation beyond the extensive margins of permanent agriculture. He builds on the notion of peasantry, where the subsistence farmers require a wide selection of natural commodities to survive and pursue their cultural activities. In the absence of markets, such commodities tend to be obtained from the forest environment or through agricultural activities of limited scope. In cases of a natural increase in population, the pressure brought to bear on the land resources results in technological intensification which, in the initial phases, involves a reduction in the rotation times of the shifting cultivation. As a result, the nutritive requirements of a household combined with the accelerated rotations explained the diversity in crop selection for most households, while the reduced areas of cultivation accounted for the magnitude of production (Walker 1999).

In the field of policy analysis and planning, much work has been done, for example, to evaluate the impact of a number of agricultural policies on regional structural changes (Happe 2004), and the impacts of free trade policies on the diffusion of innovation in agricultural regions of Chile (Berger 2001). The pioneering work by Balmann (1997) was a demonstration of the existence of a dependence on paths in the evolution of land use, which he later used to investigate the effect of reducing price support and introducing compensation payments (Balmann et al. 2002). Several studies have attempted to use ABMs to explore the likely impacts of specific real-world policies (see Weisbuch and Boudjema 1999; Deffuant et al. 2002; Sengupta et al. 2005; Janssen 2001), while others have examined the influence of generic and abstract policies on the behavior of an agent within a system (Janssen et al. 2000).

Deadman et al. (2004) presented a simulation model that explored human understanding of the spatial, social and environmental concerns related to LUCC. Based on a heuristic decision-making strategy, they utilized household characteristics, among other factors, in which the interaction of agents was effected through a labor pool. While subsistence labor demands may not always be significant, it has been reported that significant gender differences occur with respect to farm labor within households (Siqueira et al. 2002) in much the same way as population age. Although it flexed randomly on gender, the LUCITA model (see Deadman et al. 2004) did not pay particular attention to overall population age. Evans and Kelley (2004) did an analysis of scale and how it impacts on the design and implementation of LUCC ABMs at the household micro-level. The analysis revealed differences in land-use preference weights that helped to identify scale considerations in the design, development, validation and application of ABMs in LUCC analysis. In their discussion, Evans and Kelley (2004) highlight the complexities of spatial scale and computational capacity limitations, and acknowledge the non-monetary influences on decision-making.

Using ABMs to describe the decision making of land-use parcel managers and cellular automata to represent the landscape, the SLUDGE model explored the impact of distance-dependent spatial externalities and transportation costs on patterns of urban development and land use (Parker and Meretsky 2004). A similar test on the mechanisms behind the growth and spatial patterns of cities was conducted by Torrens and Alberti (2000) in order to address issues of local decision-making in determining urban sprawl. In this study, several metrics were developed to quantify the sprawl patterns. Brown et al. (2004), Loibl and Toetzer (2003), Rajan and Shibasaki (2000), Sanders et al. (1997), Dean et al. (2000), Kohler et al. (2000), Hoffmann et al. (2002), Huigen (2004), and Otter et al. (2001) have all contributed significantly to the use of ABMs by explicitly simulating human decision-making processes rather than using empirical approaches (Mathews et al. 2007).

10.5 Conclusions

Agent-based modeling is an approach that continues to receive attention in studies of many geographical phenomena. As Mathews et al. (2007) note, this is because it offers a way of incorporating the influence of human decision-making on land use in a mechanistic, formal and spatially explicit way. ABM is therefore a handy tool in developing a greater understanding of the natural world.

Empirical illustrations of observed outcomes have been shown to be a sufficient end for ABM (Epstein 1999). However, Parker et al. (2003) note that it is retrogressive to limit the potential and appropriateness of ABM to such illustrations, especially in cases where the design and implementation prospects of ABMs are very promising, with reported success in many varying fields of human significance. It has been argued that as simulation models, ABMs are limited. Firstly, they cannot be sufficiently deductive to give confidence in the outcomes from the model parameters. However, as Judd (1997) counter-argued, through sensitivity analysis, an almost complete understanding of the dynamic system under study is achievable. Secondly, ABMs are said to be sensitive to small perturbations in model parameter values at the micro-scale or a lower level, thereby providing a multitude of outcomes, but as Parker and Meretsky (2004) stated, the focus of ABMs is on the macro-scale or emergent patterns. Although there may be significant differences at the micro-level, the outcomes tend to be similar at the macro-level. While ABMs address individualism in the mechanics of system behavior, validating ABMs has proved to be a difficult task. However, Berger (2001) justified his choice by pointing out that ABMs do allow for a pragmatic treatment of data availability. He cited the exchange of information interactions between farming households, the cumulative effects of experience and the observation of neighbors' experience, and technical and financial constraints as factors that affected the diffusion of innovations and which could be explicitly defined and controlled within an ABM.

ABMs are implemented at varying levels of stakeholder involvement. Parker et al. (2003) highlighted three cases in which stakeholders were involved either right from the beginning of the modeling process, in the final stages of testing and running the model, or where models are presented as ready-made applications to policy makers. With the majority of ABMs falling into the former two categories, Mathews et al. (2007) ascribed the failure by end-users to use ABMs directly as decision support systems to a poor understanding by researchers of the actual process of decision-making and the role that decision support tools may play in this process. Several other factors are attributed to the lack of success of decision support systems, since failures at the former two levels are equally common (Mathews et al. 2007). Faced with such limitations, Stephens and Middleton (2002) stated that simulation models are probably more useful as research tools to provide insights into constraints that can later be transformed into rules-of-thumb, than as operational decision support tools. Lempert (2002) followed a similar line when he argued that much of the failure is to do with the predictive, as opposed to explanatory, approach that many modellers adopt. He suggested that model runs ought to compare the robustness, resilience and stability of alternative policies.

Agents interact either indirectly through a shared environment and/or directly with each other through markets, social networks and/or institutions. Higher-order variables such as commodity prices and population dynamics are usually expressed as emergent outcomes (Mathews et al. 2007). Moving from relatively abstract representations, ABMs have gradually progressed into an exploration of the conceptual aspects of spatially explicit systems of real-world situations (see Epstein and Axtell 1996), and all the way through to more complex representations of socio-ecological systems (see Berger and Ringler 2002; Hoffmann et al. 2002). With the addition of empirical data, recent versions of these models are now being applied to specific real-world situations (see Deadman et al. 2004). Complex environmental problems tend to be multidisciplinary, temporally dynamic, and spatially referenced. As a result, the nature of the interactions of these systems often makes it difficult to predict the outcomes for particular management actions, socio-economic conditions, or environmental processes (Deadman et al. 2004). However, recent advances in computing technology have further enhanced the use of computer-based models and analyzes that have since expanded the interest in computational approaches to the study of human geographic systems with the aim of providing meaningful solutions. ABM has tapped into these advances, and there is still plenty of room for growth and improvement.

References

Auyung SY (1998) Foundations of complex-systems theories in economics, evolutionary biology, and statistical physics. Cambridge University Press, Cambridge

Axelrod R, Cohen MD (2000) Harnessing complexity: organizational implications of a scientific frontier. Basic Book, New York

Baker WL (1989) A review of models in landscape change. Landsc Ecol 2(2):111–133

Balmann A (1997) Farm-based modelling of a region structural change: a cellular automata approach. Eur Rev Agric Econ 24:85–108

Balmann A, Happe K, Kellermann K, Kleingarn A (2002) Adjustment costs of agri-environment policy switching: an agent-based analysis of the German region Hohenlohe. In: Janssen M (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 127–157

Balzter H, Braun PW, Kohler W (1998) Cellular automata models for vegetation dynamics. Ecol Model 107(2/3):113–125

Bandini S, Manzoni S, Vizzari G (2009) Agent based modeling and simulation: an informatics perspective. J Artif Soc Soc Simulat 12(4):4. http://jasss.soc.surrey.ac.uk/12/4/4.html. Accessed 14 Apr 2011

Becu N, Perez P, Walker B, Barreteau O, Le Page C (2003) Agent-based simulation of a small catchment water management in northern Thailand: description of the Catchscape model. Ecol Model 170:319–331

Berger T (2001) Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. Agr Econ 25:245–260

Berger T, Ringler C (2002) Trade-offs, efficiency gains, and technical change: modelling water management and land use within a multiple-agent framework. Q J Int Agr 41(1/2):119–144

Bonabeau E (2002) Agent-based modelling: methods and techniques for simulating human systems. Proc Natl Acad Sci 99(3):7280–7287

Bousquet F, Le Page C (2004) Multi-agent simulations and ecosystem management: a review. Ecol Model 176(3–4):313–332

Brown DG, Page SE, Riolo R, Rand W (2004) Agent-based and analytical modelling to evaluate the effectiveness of green belts. Environ Model Softw 19:1097–1109

Buchta C, Meyer D, Ster AP, Mild A, Taudes A (2003) Technological efficiency and organizational inertia: a model of the emergence of disruption. Comput Math Organ Theor 9:127–146
Casti JL (1995) Complexication: explaining a paradoxical world through the science of surprise, 1st edn. Harper Perennial, New York
Chuvieco E (1993) Integration of linear programming and GIS for land-use modeling. Int J Geogr Inform Syst 7(1):71–83
Collier N (2000) RePast: an extensible framework for agent simulation. http://repast.sourceforge.net/docs/repast_intro_final.doc. Accessed 14 Jun 2011
Conte R, Hegselmann R, Terna P (eds) (1997) Simulating social phenomena. Springer, Berlin
Couclelis H (2001) Modeling frameworks, paradigms, and approaches. In: Clarke KC, Parks BE, Crane MP (eds) Geographic information systems and environmental modeling. Longman and Co, New York, pp 33–48
Cromley RG, Hanink DM (1999) Coupling land-use allocation models with raster GIS. J Geogr Syst 1:137–153
Dawid H, Dermietzel J (2006) How robust is the equal split norm? Responsive strategies, selection mechanisms and the need for economic interpretation of simulation parameters. Comput Econ 28:371–397
Deadman P, Robinson D, Moran E, Brondizo E (2004) Colonist household decision-making and land-use change in the Amazon Rainforest: an agent based simulation. Environ Plann Plann Des 31:693–709
Dean JS, Gumerman GJ, Epstein JM, Axtell RL, Swedlund AC, Parket MT, McCarroll S (2000) Understanding Anasazi cultural change through agent-based modelling. In: Kohler TA, Gumerman GJ (eds) Dynamics in human and primate studies: agent-based modeling of social and spatial processes. Oxford University Press, New York, pp 179–206
Deffuant G, Huet S, Bousset JP, Henriot J, Amon G, Weisbuch G (2002) Agent-based simulation of organic farming conversion in Allier department. In: Janssen M (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 158–187
Epstein JM (1999) Agent-based models and generative social science. Complexity 4(5):41–60
Epstein JM, Axtell R (1996) Growing artificial societies: social science from the bottom up. Brookings Institute, Washington
Evans TP, Kelley H (2004) Multi-scale analysis of a household-level agent-based model of land cover change. J Environ Manag 72:57–72
Geoghegan J et al (1998) Socializing the pixel and pixelizing the social in land-use and land-cover change. In: Liverman D, Moran EF, Rindfuss RR, Stern PC (eds) People and pixels: linking remote sensing and social science. National Academy Press, Washington, pp 51–69
Gilbert N, Troitzsch KG (1999) Simulation for the social scientist. Open University Press, London
Gleick J (1987) Chaos: making a new science. Viking, New York
Happe K (2004) Agricultural policies and farm structures. Agent-based modelling and application to EU-policy reform. Studies on the agricultural and food sector in central and eastern Europe, 30, IAMO. http://www.iamo.de/dok/sr_vol30.pdf. Accessed 23 June 2011
Heath BL (2010) The history, philosophy, and practice of agent-based modeling and the development of the conceptual model for a simulation diagram. Ph.D. Thesis, School of Graduate Studies, Wright State University. http://etd.ohiolink.edu/send-pdf.cgi/Heath%20Brian.pdf?wright1269176275. Accessed 8 Jun 2011
Hoffmann M, Kelley H, Evans T (2002) Simulating land-cover change in South-Central Indiana: an agent-based model of deforestation and afforestation. In: Janssen M (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 218–247
Hogeweg P (1988) Cellular automata as a paradigm for ecological modelling. Appl Math Comput 27(1):81–100
Howitt RE (1995) Positive mathematical programming. Am J Agr Econ 77(2):329–42
Huigen MGA (2004) First principles of the MameLuke multi-actor modelling framework for land use change, illustrated with a Philippine case study. J Environ Manag 72:5–21

Janssen MA (2001) An exploratory integrated model to assess management of lake eutrophication. Ecol Model 140:111–124
Janssen MA, Jager W (2000) The human actor in ecological economic models. Ecol Econ 35(3):307–310
Janssen MA, Walker BH, Langridge J, Abel N (2000) An adaptive agent model for analysing co-evolution of management and policies in a complex rangeland system. Ecol Model 131:249–268
Judd KL (1997) Computational economics and economic theory: substitutes or complements. J Econ Dyn Control 21(6):907–942
Kamusoko C, Masamu A, Bongo A, Munyaradzi M (2009) Rural sustainability under threat in Zimbabwe: simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Appl Geogr 29:435–447
Kohler TA, Kresl J, West CV, Carr E, Wilshusen RH (2000) Be there then: a modelling approach to settlement determinants and spatial efficiency among late ancestral populations of the Mesa Verde region, US Southwest. In: Kohler TA, Gumerman GJ (eds) Dynamics in human and primate studies: agent-based modeling of social and spatial processes. Oxford University Press, New York, pp 145–178
Langton CG (1989) Artificial life. In: Langton CG (ed) Artificial life. Addison-Wesley, Redwood City, pp 1–48
Lempert R (2002) Agent-based modeling as organizational and public policy simulators. Proc Natl Acad Sci 99:7195–7196
Levy S (1992) Artificial life: a report from the frontier where computers meet biology. Vintage Books, New York
Li H, Reynolds JF (1997) Modelling effects of spatial pattern, drought, and grazing on rates of rangeland degradation: a combined Markov and cellular automaton approach. In: Quattrochi DA, Goodchild MF (eds) Scale in remote sensing and GIS. Lewis Publishers, New York, pp 211–230
Liebrand WBG, Nowak A, Hegselmann R (eds) (1988) Computer modeling of social processes. SAGE Publications, London
Ligtenberg A, Wachowicz M, Bregt AK, Beulens A, Kettenis DL (2004) A design and application of a multi-agent system for simulation of multi-actor spatial planning. J Environ Manag 72(1–2):43–55
Loibl W, Toetzer T (2003) Modeling growth and densification processes in suburban regions: simulation of landscape transition with spatial agents. Environ Model Softw 18(6):553–563
Longley P, Higgs G, Martin D (1994) The predictive use of GIS to model property valuations. Int J Geogr Inform Syst 8(2):217–235
Ludeke AK, Maggio RC, Reid LM (1990) An analysis of anthropogenic deforestation using logistic regression and GIS. J Environ Manag 31:247–259
Luke S, Cioffi-Revilla C, Panait L, Sullivan K, Balan G (2005) MASON: a multiagent simulation environment. Simulation 81:517–527
Malleson N (2010) Agent-based modelling of burglary. Ph.D. Thesis, School of Geography, University of Leeds. http://www.geog.leeds.ac.uk/fileadmin/downloads/school/people/postgrads/n.malleson/thesis-final.pdf. Accessed 10 Mar 2011
Manson SM (2005) Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. Agr Ecosyst Environ 111:47–62
Mathews RB, Gilbert NG, Roach A, Polhill J, Gotts NM (2007) Agent-based land-use models: a review of applications. Landsc Ecol 22:1447–1459
Mertens B, Lambin EF (1997) Spatial modelling of deforestation in southern Cameroon. Appl Geogr 17(2):143–162
Minar N, Burkhart R, Langton C, Askenazi M (1996) The Swarm simulation system: a toolkit for building multi-agent simulations, June. Report No.: 96-06-042. Santa Fe Institute, Santa Fe
O'Sullivan D, Haklay M (2000) Agent-based models and individualism: is the world agent-based? Environ Plann 32(8):1409–1425
Otter HS, van der Veen A, de Vriend HJ (2001) ABLOoM: location behaviour, spatial patterns, and agent-based modelling. J Artif Soc Soc Simulat 4(4):2. http://www.jasss.soc.surrey.ac.uk/4/4/2.html. Accessed 1 Jun 2011

Parker DC, Meretsky V (2004) Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. Agr Ecosyst Environ 101:233–250

Parker DC, Manson SM, Janssen MA, Hoffmann M, Deadman P (2003) Multi-Agent systems for the simulation of land-use and land-cover change: a review. Ann Assoc Am Geogr 93:314–337

Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82:609–623

Rajan KS, Shibasaki R (2000) A GIS based integrated land use/cover change model to study human–land interactions. Int Arch Photogramm 13, Part B7

Rindfuss RR, Walsh SJ, Mishra V, Fox J, Dolcemascolo GP (2003) Linking household and remotely sensed data; methodological and practical problems. In: Fox J, Rindfuss RR, Walsh SJ, Mishra V (eds) People and the environment: approaches for linking household and community surveys to remote sensing and GIS. Kluwer Academic, Boston, pp 1–31

Russel S, Norvig P (1995) Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River

Sanders L, Pumain D, Mathian H, Guerin-Pace F, Bura S (1997) SIMPOP: a multi-agent system for the study of urbanism. Environ Plann Plann Des 24:287–305

Schreinemachers P, Berger T (2006) Land use decisions in developing countries and their representation in multi-agent systems. J Land Use Sci 1(1):29–44

Sengupta R, Lant C, Kraft S, Beaulieu J, Peterson W, Loftus T (2005) Modeling enrolment in the conservation reserve program by using agents within spatial decision support systems: an example from southern Illinois. Environ Plann Plann Des 32(6):821–834

Simon HA (1957) Models of man. Wiley, New York

Siqueira AD, McCracken SD, Brondizo ES, Moran EF (2002) Women in a Brazilian agricultural frontier. In: Clark G (ed) Gender at work in economic life. University Press of America, Lanham, pp 243–267

Sklar FH, Costanza R (1991) The development of dynamic spatial models for landscape ecology: a review and prognosis. In: Tuner MG, Gardner RH (eds) Quantitative methods in landscape ecology. Springer, New York, pp 239–288

Stephens W, Middleton T (2002) Why has the uptake of decision-support systems been so poor? In: Matthews RB, Stephens W (eds) Crop-soil simulation models: applications in developing countries. CAB Internationa, Wallingford, pp 129–147

Tobias R, Hofmann C (2004) Evaluation of free Java-libraries for social-scientific agent-based simulation. J Artif Soc Soc Simulat 7. http://jasss.soc.surrey.ac.uk/7/1/6.html. Accessed 3 May 2011

Torrens P, Alberti M (2000) Measuring sprawl. Working Paper 27. CASA, University College London, London

Verburg PH (2006) Simulating feedbacks in land use and land cover change models. Landsc Ecol 21:1171–1183

Verburg P, Kok K, Veldkamp T (2005) Pixels or agents? Modelling land-use and land-cover change. IHDP Update 03(2005):8–9

von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Urbana

Wada Y, Rajan KS, Shibasaki R (2007) Modelling the spatial distribution of shifting cultivation in Luangprabang, Lao PDR. Environ Plann Plann Des 34:261–278

Walker R (1999) The structure of uncultivated wilderness: land use beyond the extensive margins. J Reg Sci 39:387–410

Weinberg M, Kling CL, Wilen JE (1993) Water markets and water quality. Am J Agr Econ 75(2):278–291

Weisbuch G, Boudjema G (1999) Dynamical aspects in the adoption of agri-environmental measures. Adv Complex Syst 2:11–36

Weiss G (ed) (1999) Multi-agent systems: a modern approach to distributed artificial intelligence. MIT Press, Cambridge

Weyns D, Omicini A, Odell J (2007) Environment as a first class abstraction in multiagent systems. Auton Agent Multi-Agent Syst 14(1):5–30

Chapter 11
Analytic Hierarchy Process in Geospatial Analysis

Ronald C. Estoque

11.1 Introduction

Owing to the increasing demand for land, forests, waterways, and other depleting resources brought about by increasing population growth, there is an urgent need for a scheme that can promote their sustainable utilization. The best use of these resources must be selected so that they remain available to the generations to come. Natural disasters like earthquakes, landslides, and floods have been major concerns in many countries. This makes it necessary for planners to design sound risk-management contingencies to prepare for such disasters. Most of the time, however, decision makers have different and conflicting priorities, concerns, knowledge, and expertise in dealing with these problems. This reality complicates the decision-making process on how a particular resource should be utilized, and on how an analysis of the susceptibility of a particular area to a certain disaster risk hazard should be carried out. In recognition of the complexity, magnitude, and importance of these problems, a decision-making technique that is responsive, transparent, and acceptable to the decision makers and other stakeholders is needed. Multi-criteria decision making (MCDM), or multi-criteria decision analysis (MCDA), is a decision-making technique that helps decision makers who are confronted with conflicting priorities to come up with an acceptable decision using a transparent decision-making process. It has been one of the fastest growing problem-resolving approaches in the past decades (Triantaphyllou 2000).

Many problems associated with geospatial considerations have been resolved using geographic information systems-based MCDA (GIS-based MCDA) (Malczewski 2006). GIS is recognized as a decision support system that involves

R.C. Estoque (✉)
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Don Mariano Marcos Memorial State University, Bacnotan, La Union, Philippines
e-mail: purplebee80@yahoo.co.uk

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_11,

the integration of spatially referenced data in a problem-solving environment (Cowen 1988). It provides supporting techniques and procedures that are vital in analyzing decision problems (Malczewski 2006). MCDA, on the other hand, provides the necessary techniques and procedures for structuring decision problems, and designing, evaluating, and prioritizing alternative decisions. Thus, GIS-based MCDA can be regarded as a process that combines geographical data (spatial data) and value judgments or the decision maker's preferences (aspatial data) to obtain information for decision making (Malczewski 2004, 2006).

Within the umbrella of GIS-based MCDA, three dichotomies exist (Malczewski 2006), namely (a) multi-attribute decision analysis (MADA) versus multi-objective decision analysis (MODA), (b) individual versus group decision making, and (c) decisions under certainty versus decisions under uncertainty (i.e., probabilistic and fuzzy decision making). The MADA methods are "data-oriented," and are also referred to as discrete methods because they assume that the number of alternatives (plans) is given explicitly, while MODA methods are "mathematical programming model-oriented" and the alternatives must be generated (they are identified by solving a multi-objective mathematical programming problem) (Malczewski 2004). Some of the GIS-based MCDA procedures or decision rules, particularly for MODA, include multi-objective programming algorithms (linear-integer programming), heuristic search/evolutionary/genetic algorithms, and goal programming/reference point algorithms. On the other hand, some of the MADA methods that have already been investigated and applied include the weighted linear combination (WLC) or simple additive weighting, Boolean overlay, ideal/reference point methods [e.g., technique for order preference by similarity to ideal solution (TOPSIS) or multi-objective land allocation (MOLA)], concordance analysis, outranking methods [e.g., elimination and choice translating reality (ELECTRE) or preference ranking organization method for enrichment evaluation (PROMETHEE), and the analytic hierarchy process (AHP)] (Malczewski 2004, 2006).

While discussing the applications of all the above-mentioned methods is beyond the scope of this chapter, Triantaphyllou (2000) provides a comprehensive comparative study of most of the MADA methods mentioned. There is also a well-established body of literature on the much wider perspective of GIS-based MCDA (Malczewski 2006), including the works of Diamond and Wright (1988), Janssen and Rietveld (1990), Carver (1991), Church et al. (1992), Banai (1993), Pereira and Duckstein (1993), Eastman et al. (1995), Heywood et al. (1995), Jankowski (1995), Laaribi et al. (1996), Malczewski (1999), Thill (1999), Chakhar and Martel (2003), and Feick and Hall (2004). Malczewski (2004, 2006) wrote a critical overview on GIS-based land suitability analysis, and surveyed the literature on GIS-based MCDA. Mendoza and Martins (2006) also published a critical review on the methods and new modeling paradigms that incorporate MCDA in natural resource management.

This chapter focuses on a particular decision support tool, the AHP method. Interest in using this method has been growing continuously over the past two decades, and its scope of application has been expanding, especially in the field of geospatial analysis. This chapter outlines the basic principles of AHP, and reviews its historical development and application as a support tool for GIS-based MCDA. The remainder of this chapter is organized as follows: (a) an overview on how AHP works, (b) AHP methodological development, including the historical development of its application as a weighting method for GIS-based MCDA, (c) a review of previous empirical studies, and (d) concluding remarks and future prospects.

11.2 An Overview on How AHP Works

The AHP (Saaty 1980) is a multiple criteria decision-making tool that has been used in many applications related to decision making. The effectiveness of AHP in evaluating problems involving multiple and diverse criteria and the measurement of trade-offs—sometimes using limited available data (Banai 1989)—has led to its worldwide recognition across different fields of application. It has been applied to a wide range of applications from selecting among competing alternatives in a multi-objective environment, allocation of scarce resources, to forecasting (Forman and Gass 2001). More particularly, AHP has been successfully applied to many complex planning, resource allocation, and priority-setting problems in business, energy, health, marketing, transportation, natural resources, and environmental sciences (Schmoldt et al. 2001). Over the years, many researchers, scientists, and decision makers in various fields have implemented AHP. In a recent review, Vaidya and Kumar (2006) presented ten different thematic areas where AHP has been applied. These are selection, evaluation, benefit–cost analysis, allocations, planning and development, priority and ranking, decision making, forecasting, medicine and related fields, and quality function deployment (QFD).

More specifically, AHP is a decision-making approach based on the genuine ability of people to make critical decisions (Saaty 1994a). It allows the active participation of decision makers in exploring all possible options in order to fully understand the underlying problems before reaching an agreement or arriving at a decision (Estoque and Murayama 2010; Yalcin 2008). Its fundamental purpose is to judge the given alternatives for a particular goal by developing priorities for these alternatives and for the selected criteria (Saaty 2001). A pair-wise comparison technique is used to derive the priorities for the criteria in terms of their importance in achieving the goal. Similarly, the priorities for the alternatives (i.e., the competing choices under consideration) are derived in pair-wise comparisons in terms of their performance against each criterion. Generally, AHP is based on three principles: decomposition, comparative judgment, and synthesis of priorities (Malczewski 1999; Saaty 1980, 2008a). Its whole process includes five fundamental steps, as presented below. Teknomo (2006) and Coyle (2004) provide open access comprehensive tutorials on the basic implementation of these steps in the AHP.

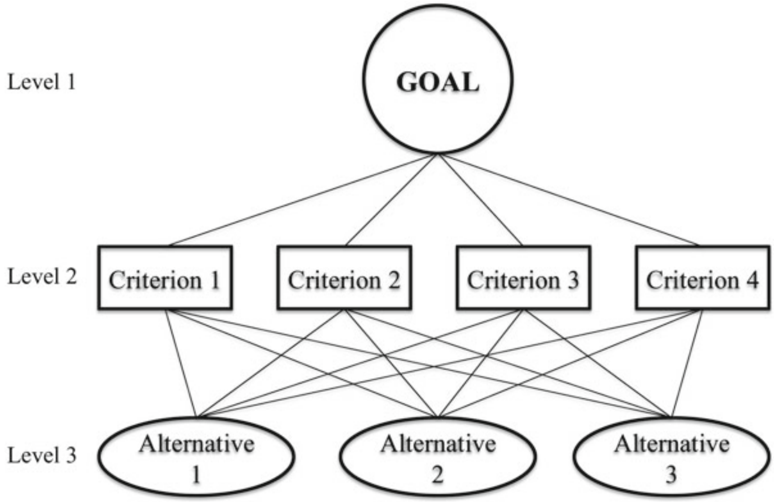


Fig. 11.1 The general structure of AHP for multi-criteria decision making (after Zahedi 1986, p. 97). The goal is to choose among the competing alternatives 1, 2, and 3 on the basis of a ranking score when judged individually against criteria 1, 2, 3, and 4

11.2.1 Step 1: Modeling the Problem

The very first step includes stating the problem, broadening the objectives of the problem by considering all actors, objectives and corresponding outcomes, and the identification of decision elements such as alternatives and criteria or decision rules. The decision elements are set up into a hierarchy of interrelated decision elements constituting the goal, criteria, sub-criteria, and alternatives (Johnson 1980; Vaidya and Kumar 2006). This step has been thought to be the most important aspect of AHP (Zahedi 1986). At the topmost position of the hierarchy is the overall goal (i.e., level 1), such as the goal of selecting the best alternative. The next lower level (i.e., level 2) of the hierarchy includes the decision rules or criteria that contribute to the attainment of the overall goal. This level can be expanded depending on how much detail is considered for each decision rule or criterion. The lowest level (i.e., level 3) contains the alternative decisions from which the decision analyst/maker will select. A simplified general structure of the AHP is presented in Fig. 11.1.

11.2.2 Step 2: Determining Priorities Among the Decision Elements of the Hierarchy

This step involves the gathering of ratings for each of the criteria and alternatives using a pair-wise comparison technique and the rating scale of relative importance.

Table 11.1 The fundamental scale of absolute numbers (Saaty 2008b, p. 86)

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities (elements) contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgment slightly favor one activity (element) over another
4	Moderate plus	
5	Strong importance	Experience and judgment strongly favor one activity (element) over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity (element) is favored very strongly over another; its dominance is demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favoring one activity (element) over another is of the highest possible order of affirmation
Reciprocals of above	If activity (element) *i* has one of the above non-zero numbers assigned to it when compared with activity (element) *j*, then *j* has the reciprocal value when compared with *i*	A reasonable assumption
1.1–1.9	If the activities (elements) are very close	May be difficult to assign the best value, but when compared with other contrasting activities (elements) the size of the small numbers would not be too noticeable, yet they can still indicate the relative importance of the activities (elements)

This step invokes the participation of experts and/or stakeholders in determining the relative importance of one criterion or alternative over another through a pair-wise comparison method presented in a matrix (see Saaty 2008a, 2008b; Teknomo 2006; Coyle 2004 for examples). The number of comparisons for the decision elements in a particular level is derived using (11.1) (Teknomo 2006; Vaidya and Kumar 2006). Each comparison (e.g., Criteria 1 vs. Criteria 2 or Alternative 1 vs. Alternative 2) is rated by a group of experts using the scale developed by Saaty (1980, 2008b) for a pair-wise comparison technique (Table 11.1). To incorporate a group consensus, the process generally includes a questionnaire for comparing all elements and a geometric mean to arrive at a final solution (Vaidya and Kumar 2006).

$$\textit{Number of comparisons} = \frac{n(n-1)}{2} \tag{11.1}$$

where n is the number of elements (i.e., criteria or alternatives).

11.2.3 Step 3: Deriving the Overall Relative Weights of the Decision Elements

In this step, the relative importance of the criteria, as far as the attainment of the goal is concerned, and the relative importance of the alternatives with respect to the criteria are determined after a pair-wise comparison matrix for the criteria and for the alternatives have been prepared (Step 2). This is done by: (1) calculating the normalized values for each criterion and alternative, and (2) determining the normalized principal eigenvectors or priority vectors (herein also referred to as relative weights). In calculating the normalized values for each criterion and alternative in their respective matrices, the value for each cell is divided by its column total. This process produces a column total of 1 for each criterion and alternative. The relative weights are then calculated by averaging the rows of each matrix. The resulting values give the relative weights of the criteria with respect to the goal, and the relative weights of the alternatives with respect to the criteria. The overall relative weights of the alternatives are determined by calculating the linear combination of the product between the relative weight of each criterion and the relative weight of the alternative for that criterion (Coyle 2004; Forman and Gass 2001; Saaty 1980, 2008b; Teknomo 2006; Zahedi 1986). If the expert judgments are consistent (see Steps 4 and 5), the decision makers then select the best choice based on the overall relative weights of the alternatives.

11.2.4 Steps 4 and 5: Verifying the Consistency of Judgments and Making Conclusions Based on the Results

These steps are necessary to determine the consistency of the evaluation by calculating the consistency ratio (CR) before a decision is made. If the problem under consideration was aimed at selecting the best alternative, the CRs for all the matrices (i.e., for the criteria and the alternatives) are calculated first before the overall relative weights of the alternatives are computed. The CR for a particular matrix is determined using (11.2) and (11.3). Saaty (1980) suggests that if the ratio exceeds 0.1, the set of judgments may be too inconsistent to be reliable. Thus, a CR below 0.1 or 10% is acceptable. When the evaluation is inconsistent, the procedure is repeated until the CR is within the desired range. Decision makers then reach a conclusion based on the results.

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{11.2}$$

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1} \tag{11.3}$$

where CR is the consistency ratio, CI is the consistency index, RI is the random consistency index, $\lambda_{\max}$ is the principal eigenvalue or summation of the products

Table 11.2 The random consistency index (RI) values (from Saaty 1980, p. 21)

n	1	2	3	4	5	6	7	8	9	10	11	12
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48

between each element of the relative weights and their corresponding column total in a particular matrix, and n is the number of elements (i.e., the number of criteria or alternatives). The RI developed by Saaty (1980) is presented in Table 11.2.

11.3 Methodological Development

11.3.1 Historical Development of AHP

AHP was developed and introduced in the 1970s by Thomas L. Saaty, an American mathematician and professor of the University of Pittsburgh, Pennsylvania, USA (Schmoldt et al. 2001; Yang and Shi 2002). In the late 1960s, Saaty, then an operations research pioneer, directed research projects for the Arms Control and Disarmament Agency at the US Department of State (Forman and Gass 2001). The development of AHP was motivated by his disappointment with his team's output in their quest for a practical systematic approach to priority setting and decision making to support arms-reduction negotiations in Geneva between the USA and the Soviet Union (Forman and Gass 2001; Nekhay et al. 2009). Saaty (2001) explained that through AHP, a decision maker would be able to construct hierarchies or feedback networks that describe the environment structure of the decision for a particular complex problem. Once done, the decision maker would then be able to make judgments or perform measurements on pairs of decision elements with respect to a controlling element or the goal under consideration. This is in order to derive ratio scales or priority values that are then synthesized throughout the structure to select the best alternative. Saaty's earlier case applications of AHP range from the choice of a school for his son, through to the planning of transportation systems for Sudan (Coyle 2004). In accordance with Saaty (1980, 1988, 1994a, 1994b), Zahedi (1986) and Malczewski (1999), Forman and Gass (2001) described AHP as a methodology for *structuring, measurement, and synthesis* towards solving any given complex problem.

Structuring in AHP involves modeling a specific complex problem in such a way that the interrelated decision elements (i.e., criteria and alternatives) are structured into a hierarchy, as discussed in the previous section. *An AHP measurement* includes comparisons of the decision elements in a pair-wise manner based on the preference of the decision makers and experts. AHP is fundamentally based on a clearly outlined mathematical structure of matrices coupled with their right-eigenvector's ability to give true or approximate weights (Saaty 1980, 1994a). In AHP, pair-wise comparison is a major step to determine the relative importance of the decision elements, to determine the weight of each criterion or factor, and to provide a rating for alternatives based on qualitative factors (Yang and Shi 2002). This step particularly focuses on the evaluation of two elements at a time, and therefore the decision makers and experts are under less pressure in giving their own relative preference information. The numerical scale developed by Saaty (1980, 2008b) (Table 11.1) is used to give numerical ratings that correspond to the experts' judgments on the relative importance of the decision elements. The *synthesis* function is particularly concerned with putting the pieces together to get the final result of the analysis. AHP translates individual preferences or judgments into ratio-scale weights that are combined into linear additive weights for the associated alternatives (Forman and Gass 2001). The derived weights are used to rank the given alternatives. In this way, AHP aids the decision makers in choosing the best alternative. In fact, the determination of the priority weights is a very important step after pair-wise comparisons have been done. In calculating these weights, three methods have been suggested in the research literature, these include calculating the normalized eigenvalues, logarithmic least squares, and least squares (Yang and Shi 2002). Use of the normalized eigenvalues is recommended when the data are not entirely consistent, although it has been proven that these methods produce the same solutions in terms of the consistency of the result (Saaty 1988).

Forman and Gass (2001) summarize the *measurement* and *synthesis* functions of AHP into what they called three commonly agreed decision-making steps: (1) given $i=1,\ldots,m$ objectives/criteria, determine their respective weights w_i; (2) for each objective/criterion i, compare the $j=1,\ldots,n$ alternatives and determine their weights w_{ij} with respect to each objective/criterion i; (3) determine the final (global) alternative weights (priorities) W_j with respect to all the objectives/criteria by $W_j=w_{1j}w_1+w_{2j}w_2+\cdots+w_{mj}w_m$. W_j orders the alternatives, with the most preferred alternative having the largest W_j. As far as pair-wise comparisons are concerned, Saaty (1980) suggests that the number of elements at each level be limited to a maximum of nine, since each level entails pair-wise comparisons of its elements. Zahedi (1986) notes that although this is a constraint, it is not a necessary condition of the method and it has not always been adhered to in all applications.

From the time it was developed and introduced, AHP has been gaining popularity among decision makers, researchers, planners, and academicians. Its validity as a decision tool, including the confidence that the users are expressing in its ability to aid in MCDA situations, is supported by the many hundreds (now probably thousands) of diverse applications in which AHP results have been accepted and used by the decision makers concerned (Saaty 1994b). Quoted below is a brief description by Forman and Gass (2001, p. 470) in their exposition on AHP:

> There is ample evidence that the power and simplicity of AHP has led to widespread acceptance and usage in the United States as well as throughout the world. In addition to Expert Choice, there have been several other successful commercial implementations of the AHP, one with financial backing of the Canadian Government. Many of the world's leading information technology companies now use AHP in the form of decision models provided by the Gartner Group's [(http://www.gartner.com)] Decision Drivers [(http://www.decisiondrivers.com)]. The American Society for Testing and Materials (ASTM) has adopted AHP as standard practice for multi-attribute decision analysis of investments related to buildings and building systems [(ASTM Designation E: 1765-95 "Standard Practice for Applying Analytical Hierarchy Process (AHP) to Multiattribute Decision Analysis of Investments Related to Buildings and Building Systems")]. The AHP process is taught in numerous Universities and used extensively in organizations such as the Central Intelligence Agency that have carefully investigated AHP's theoretical underpinnings.

These remarks relate mainly to the application of AHP to non-spatial problems, focusing on determining the best alternative in a complex decision-making environment. However, its function has now been extended to the vast area of spatial analysis. Geospatial analysis is one of the fast-growing research fields where AHP has been integrated as part of the methodology and analysis.

11.3.2 Using AHP as a Weighting Method for GIS-Based MCDA

The basic rationale for introducing GIS into the broad perspective of MCDA is to add spatial dimensions to the problem-solving process (Malczewski 1999). GIS, as a technology, has evolved through three broad application fields: (1) its use as an information database to coordinate and access geographic data; (2) its use as an analytical tool to specify logical and mathematical relationships among map layers to yield new derivative maps; (3) its use as a decision support system to decide how to act upon analyzes produced (Eastman et al. 1995). It is in the second and third broad areas of GIS applications where AHP has been largely integrated as a criteria/alternatives weighting and prioritization method.

One of the earlier reports that dealt with the possible development of the integration of AHP with GIS to aid in the decision-making process for complex problems was the review of AHP as a new method for site suitability analysis by Banai (1989). Rao et al. (1991) also published a paper entitled "Weighted index model for suitability assessment – a GIS approach." According to Eastman et al. (1995), although the procedure was developed outside GIS software using a variety of analytical resources, the work of Rao et al. (1991) was the first attempt to integrate AHP with GIS. Subsequent work in line with the integration of AHP and GIS included: participatory procedures for multi-criteria evaluation in GIS (Eastman et al. 1992); a procedure for multi-objective decision making in GIS under conditions of competing objectives (Eastman et al. 1993); the design of land-use plans considering multiple objectives and participatory approaches to planning and decision making (Hutchinson and Toledano 1993); a study of habitat suitability for the Mount Graham red squirrel (Pereira and Duckstein 1993); the integration of GIS-based suitability analysis and multi-criteria evaluation in a spatial decision support system for route selection (Jankowski and Richard 1994); weighting land suitability factors by the prioritization for land-use suitability method (Xiang and Whitley 1994); raster procedures for multi-criteria/multi-objective decisions (Eastman et al. 1995); integrating GIS and MCDM methods (Jankowski 1995); a demonstration of landfill site selection (Siddiqui et al. 1996); and the generation of habitat suitability indices for the desert tortoise (Mendoza 1997).

Focusing on the methodological development of a raster procedure for multi-criteria/multi-objective decisions, Eastman et al. (1995) proposed a framework (Fig. 11.2, left) in which AHP is used as a weighting method. The framework is basically divided into two main parts: multi-criteria evaluation (MCE) and multi-objective allocation. With a single objective (e.g., finding the best site for a housing

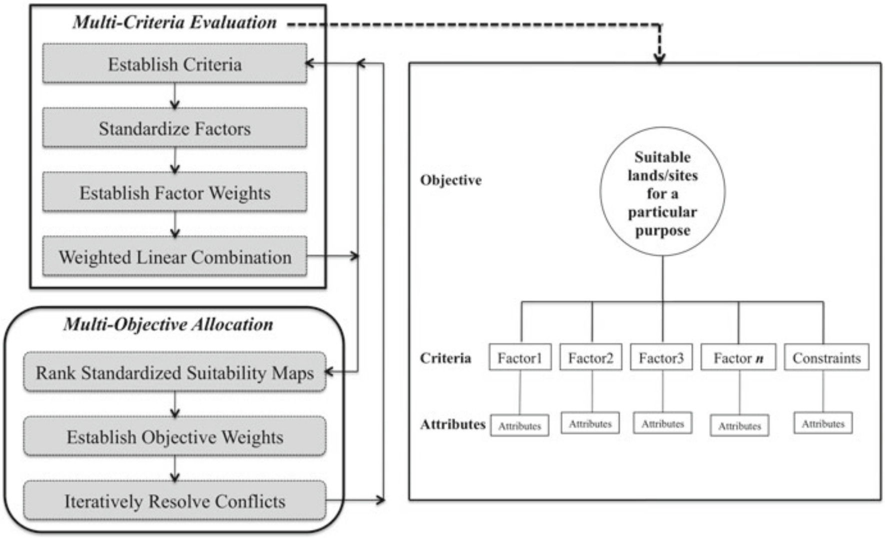


Fig. 11.2 (*Left*) A flow diagram of the multi-criteria/multi-objective decision-making process (redrawn from Eastman et al. 1995, p. 544). (*Right*) A generalized multi-criteria evaluation structure based on Table 11.3. Depending on the complexity of the problem under consideration, other elements such as sub-criteria are usually incorporated

project), the decision maker(s) can focus on relevant criteria (factors and constraints) with measurable attributes, and thus corresponding techniques are often called MCE, or multiple-attribute decision analysis/making (Fig. 11.2, right) (Jankowski 1995; Malczewski 1999). In MCE, standardization is necessary after the criteria have been identified because factors normally have different characteristics, e.g., units. This process allows the comparison of different factors, which is vital in the analysis. All factors need to be expressed in a common measurement scale before AHP weights are applied in the aggregation process (Eastman et al. 1995). Voogd (1983) reviewed a variety of procedures for standardization. Although, other scientists/researchers prefer simple reclassification and normalization as their standardization methods, others use fuzzy membership functions in IDRISI (Eastman 1999, 2006, 2009) (see Table 11.3). In a suitability analysis, the fuzzy concept is used to give each location a value representing its degree of suitability (Eastman 2006), especially if suitability values or classes do not have clearly defined boundaries. Standardized scales can be 0–1, 0–99, 0–255, etc. (Eastman et al. 1995; Eastman 2006).

WLC (weighted linear combination) (Voogd 1983), also known as simple additive weighting, is an approach that multiplies standardized or normalized criteria scores by the relative criteria weights for a particular objective, or for each alternative, followed by a summation of the results as in (11.4) (Carver 1991; Eastman, et al. 1995; Geldermann and Rentz 2007; Nyerges and Jankowski 2010; Sugumaran and Bakker 2007). Although different multi-criteria procedures can be implemented in the GIS platform, WLC in combination with AHP is the most straightforward and

Table 11.3 List of recently published studies and their details

Reference	Country of application	Subject/purpose of the study	Number/Name of criteria or factors	Criteria or factors standardization method	Number/Expertise or specialization of the decision makers or experts participated in the AHP process	Experts' ratings aggregation method/ Remarks
Estoque and Murayama (2010)	Philippines	Suitability analysis for bee-keeping sites	4/nectar and pollen source (land use/cover), distance to river, distance to road, elevation Plus a set of constraints	Fuzzy membership functions	4/beekeeping, forestry, agriculture, spatial analysis	Consensus
Fernandez and Lutz (2010)	Argentina	Urban flood hazard zoning	5/distance to the discharge channels, elevation, slope, depth to ground-water table, cover type	Re-classed and AHP weight was calculated per class	The paper mentioned, "During the analysis, weight values were assigned to the layers and rank values to the classes of each layer according to their importance in the case study of area floods."	Not applicable Not discussed in detail if experts were involved
Guiqin et al. (2009)	China	Landfill site selection	13/man and animal habitat, rivers, lakes, ground water, distance from airfields, agricultural land, forest land, special land, slope, altitude, price of land, distance from waste production centers, distance from roads	Re-classed using 1–5 rating scale	The paper mentioned, "Decision-makers evaluated the importance of pairs of grouped elements in terms of their contribution to the higher hierarchy."	The number and expertise of experts were not mentioned Aggregation method was not discussed

(continued)

Table 11.3 (continued)

Reference	Country of application	Subject/purpose of the study	Number/Name of criteria or factors	Criteria or factors standardization method	Number/Expertise or specialization of the decision makers or experts participated in the AHP process	Experts' ratings aggregation method/ Remarks
Hossain et al. (2009)	Bangladesh	Identification and prioritization of appropriate urban water bodies for carp farming	14/water temperature, water pH, dissolved oxygen, water transparency, existing utilization, slope, soil texture, soil pH, organic matter, distance to road, distance to electricity, distance to local market, distance to fry sources, labor availability	Re-classed using 1–3 rating scale	3/carp farming, however, their opinions were ignored; the evaluation of the first author was adopted	Not applicable (priority vectors were computed on the individual expert's ratings but results were later ignored)
Kamp et al. (2008)	Pakistan	Landslide susceptibility mapping	9/aspect, elevation, faults, lithology, land cover, rivers, roads, slope, tributaries	Standardization method was not discussed	The paper mentioned, "AHP considers only a one-level weighting system developed by collecting expert opinions, in this case our experience obtained during the fieldwork in November 2005."	Not applicable (maybe a consensus evaluation among the authors was used)
Khoi and Murayama (2010)	Vietnam	Delineation of suitable cropland areas	9/elevation, slope, distance to water, soil organic matter, soil depth, soil pH, soil texture, distance to road, distance to the park boundary	Fuzzy membership functions	12/agronomy, forestry, soils	Consensus

Moeinaddini et al. (2010)	Iran	Municipal solid waste landfill site selection	5/physical characteristics of the land, buffer and distances, visibility, sensitive ecosystem, land use and land cover	Fuzzy membership functions	15/expertise not mentioned	Geometric mean
Nekhay et al. (2009)	Spain	Suitability analysis of olive plantations for wildlife habitat restoration	7/water bodies, natural vegetation, natural park area, high electricity power lines, agricultural plots, roads, urban areas Four constraints	Re-classed using 1–9 rating scale and normalized to 0–1 scale	6/ecology, focal species	Geometric mean
Radiarta et al. (2008)	Japan	Identification of suitable sites for Japanese scallop cultivation	7/sea temperature, chlorophyll, suspended sediment, bathymetry, distance to towns, distance to piers, distance to land-based facilities	Suitability scores (%) were re-classed using 1–8 rating scale	The paper mentioned, "Criteria were rated according to literature reviews and experts' opinions based on their relative importance using the pairwise comparison method."	The number and expertise of experts were not mentioned Aggregation method was not discussed
Samani et al. (2010)	Iran	Planning road network in mountain forests	8/slope, elevation, aspect, geology, soil, volume, type, hydrology	Re-classed using 1–9 rating scale	Number not mentioned/ forest engineering	Arithmetic mean
Sener et al. (2010)	Turkey	Landfill site selection	9/distance from road, height, slope, aspects, distance from settlements, distance from surface water, distance from protected areas, geology/ hydrogeology, land use	Re-classed and AHP weight was calculated per class	There was no experts' participation mentioned	There was no experts' participation mentioned

(continued)

Table 11.3 (continued)

Reference	Country of application	Subject/purpose of the study	Number/Name of criteria or factors	Criteria or factors standardization method	Number/Expertise or specialization of the decision makers or experts participated in the AHP process	Experts' ratings aggregation method/ Remarks
Suárez-Vega et al. (2011)	Spain	Selection of location for establishing a new hypermarket	5/huff capture, land use, distance to the main roads, distance to the trading ports, slope of the terrain	Normalized using 0–100 rating scale	1/hypermarket (Canarian market)	Not applicable (one of the authors served as expert)
Thapa and Murayama (2008)	Vietnam	Land evaluation for peri-urban agriculture	5/soil suitability, road accessibility, land-use suitability, water accessibility, market accessibility	Re-classed to qualitative scale (least–most suitable)	12/agricultural farming and planning, marketing, community/public administration	Arithmetic mean (the consistent matrices (9 out of 12) were combined and used to prepare the final matrix)
Tudes and Yigiter (2010)	Turkey	Establishing land-use planning model based on geo-environmental criteria	8/slope, elevation, depth of ground waters, earthquake risk, surface geology, bearing power, agricultural land quality, land use	Re-classed using 0–4 rating scale	The paper mentioned, "In this study, the judgment related to the influence of each geo-environmental criterion on each land use category was made by consulting experts."	The number and expertise of experts were not mentioned Aggregation method was not discussed

Vizarri (2011)	Italy	Determining potential landscape quality	14/protected geological sites, sites of community importance, rivers, natural parks, ecological network, archeological sites, historical centers, religious architecture, military architecture, villas, typical production, traditional events, panoramic viewpoints, symbolic elements	Standardized through the calculation of the "quality of the landscape element" (output value ranges from 1 to 100)	The paper mentioned, "With the aid of experts, layer weightings were determined using three different pairwise comparison matrices [major categories of the factors]."	The number and expertise of experts were not mentioned Aggregation method was not discussed
Wu and Chen (2009)	Taiwan	Determining landslide susceptibility	7/rainfall, geology, slope, vegetation, soil moisture, road development, historic landslides	Re-classed using 1–10 rating scale	75/geology	Arithmetic mean (of the relative weights of the factors from the consistent matrices—not mentioned how many were consistent)
Yalcin (2008)	Turkey	Landslide susceptibility mapping	7/lithology-weathering, land-use, slope, aspect, distance to stream, drainage density, distance to road	Re-classed and AHP weight was calculated per class	There was no experts' participation mentioned	There was no experts' participation mentioned
Ying et al. (2007)	China	Synthetic evaluation of eco-environment quality	4/natural environment, disaster, environment pollution, social economy	Standardized using AHP (0–1 scale; magnified 100 times during the analysis)	The paper mentioned, "In this research, we invited experts with eco-environmental backgrounds to give the relative importance of each factor, ..."	The number of experts was not mentioned Aggregation method was not discussed

most frequently used (Eastman, et al. 1993; Malczewski 2004). For example, in a landslide susceptibility mapping, WLC combines thematic maps of factors or parameters controlling landslides after the standardized scores of the classes of all the parameters (primary-level weight) and the relative weights of the parameters (secondary-level weight) have been applied (Wu and Chen 2009; Yalcin 2008). However, in cases where there are constraints, the product of the constraints is further multiplied with the calculated suitability (Eastman et al. 1995) (11.5).

$$\mathrm{LSI} = \sum w_i x_i \tag{11.4}$$

$$\mathrm{LSI} = \sum w_i x_i * \prod c_j \tag{11.5}$$

where LSI is the landslide susceptibility index, w_i is the weight of factor i, x_i is the standardized value or score of factor i, c_j is the criterion score of constraint j, and Π is a product symbol (of the constraints).

When the MCDA process involves multiple objectives, the procedures outlined in the lower part of Fig. 11.2, left, are recommended. In multiple-objective problems, it is important to consider whether the objectives are complementary or in conflict (e.g., simultaneously allocating a piece of land for housing and a park), and to group the criteria by objectives (Eastman et al. 1995; Malczewski 2004). A procedure called MOLA has been developed to undertake a compromise solution to a multi-objective problem (Eastman et al. 1995; Eastman 2006, 2009). Procedurally, it first asks for the names of the objectives and their relative weights. It then asks for the names of the ranked suitability maps for each objective and the areas that should be allocated. It then iteratively reclassifies the ranked suitability maps to perform a first-stage allocation, checks for conflicts, and then allocates conflicts based on a minimum-distance-to-ideal-point rule using the weighted ranks (Eastman et al. 1995).

11.4 Review of Previous Empirical Studies

11.4.1 Previous Reviews

Among the prominent reviews on MCDA are a survey of literature on GIS-based MCDA by Malczewski (2006), a critical review on MCDA in natural resource management by Mendoza and Martins (2006), and a critical overview of GIS-based land-suitability analysis by Malczewski (2004). More focused reviews on the applications of AHP as a *stand-alone method* can be traced back to its introduction in the late 1970s. In his paper "The AHP – a survey of the method and its applications," Zahedi (1986) cited around 100 papers related to AHP. The topics range from methodology development, conflict resolution, arms control and world influence, health, marketing, and budget allocation, to interregional migration patterns and many more. In their paper "The AHP – an exposition," Forman and Gass (2001) categorize the diverse applications of AHP according to choice, prioritization/evaluation, resource allocation, bench-marking, quality management, public policy, health care, and strategic planning. More recently, in their paper "Analytic hierarchy process: an overview of applications," Vaidya and Kumar (2006) identify ten broad thematic areas where AHP has been applied as a decision support tool. However, despite these themes being broad, the review paper is partly dedicated to AHP applications in combination with finance.

11.4.2 Recent AHP Applications in GIS-Based MCDA

This section adds to the existing reviews, focusing more on the recent applications of AHP as a decision support tool or weighting method for GIS-based MCDA. It highlights how AHP and GIS-based MCDA work as an integrated method in spatial-based decision problems, with information including country of application, subject/purpose of the study, number/name of criteria or factors, criteria standardization method, number/expertise or specialization of the decision makers or experts participated in the AHP process, and experts' ratings aggregation method.

Table 11.3 summarizes the details of 18 recently published papers reporting studies conducted in 13 different countries. All the studies were implemented within the MCE platform (see Fig. 11.2, right, for the generalized MCE structure). The number of criteria or factors involved in the analyzes ranges from 4 to 14. However, studies with a large number of criteria had these criteria grouped into more general categories (e.g., Moeinaddini et al. 2010; Ying et al. 2007). Fuzzy membership functions, reclassification, normalization, and AHP methods have been used to standardize the diverse criteria. Except for the paper of Kamp et al. (2008), all the papers reviewed follow the WLC procedure in the overall aggregation process of the criteria or factors involved.

The number of experts involved in the AHP pair-wise comparison method varies from none to 75. There are papers that claim the participation of experts, but do not mention their number or their respective areas of expertise (e.g., Guiqin et al. 2009; Radiarta et al. 2008). Moreover, there are also papers that do not mention any participation of experts in the process (e.g., Fernandez and Lutz 2010; Sener et al. 2010; Yalcin 2008). Kamp et al. (2008) and Suárez-Vega et al. (2011) did not seek the opinions of experts at all; the former used their experience during fieldwork as their guide in the pair-wise comparison method, while one of the authors of the latter served as an expert. In the method of aggregating the experts' ratings in the pair-wise comparison method for the studies that consulted more than one expert, two papers use "consensus" (i.e., Estoque and Murayama 2010; Khoi and Murayama 2010), three use arithmetic mean (i.e., Samani et al. 2010; Thapa and Murayama 2008; Wu and Chen 2009), and two use geometric mean (i.e., Moeinaddini et al. 2010; Nekhay et al. 2009), while most of the rest do not discuss how the evaluators' ratings were aggregated.

Fernandez and Lutz (2010) observe that pair-wise judgments are made based on the best available information, and the decision-maker's knowledge and experience.

Likewise, Yalcin (2008) mentions that AHP allows the active participation of decision makers in reaching an agreement, and gives them a rational basis for decision making. This paper also concludes that when field conditions and characteristics are correctly determined by good expertise, the AHP approach gives better results. However, despite these claims in both papers, there is no further elucidation on whether decision makers or experts were consulted and participated in the AHP process. On the other hand, Guiqin et al. (2009), Radiarta et al. (2008), Tudes and Yigiter (2010), and Ying et al. (2007) state that decision makers/experts were consulted and that their opinions were used for rating the criteria. However, their papers do not disclose whether the ratings were consensual or if they were individually expressed by the experts, as is the case in some other papers (see Estoque and Murayama 2010; Hossain et al. 2009; Khoi and Murayama 2010; Nekhay et al. 2009; Thapa and Murayama 2008). Hossain et al. (2009) consulted three experts, but their views were later ignored in favor of the first author's own evaluations, which were slightly more consistent. Such a move notwithstanding, the experts' opinions should always be among the different viewpoints taken into consideration to obtain a balanced and informed basis for the final decision making. It is also useful to capture the different opinions of experts with diverse expertise and backgrounds. In contrast, Thapa and Murayama (2008), Wu and Chen (2009), and Samani et al. (2010) combine the individual expert's evaluations by calculating the arithmetic means. Thapa and Murayama (2008) consulted 12 experts; of this number, nine were found to be consistent. In order to prepare the final matrix and calculate the final relative weights of the factors under consideration, the consistent evaluations were combined by calculating the (arithmetic) mean. In the case of Wu and Chen (2009), the relative weights of the factors under consideration were derived from the (arithmetic) means of the largest eigenvalues of the pair-wise comparison matrix, which was established from the results of 75 expert questionnaires with a CR of less than 0.1. However, while it was mentioned that Samani et al. (2010) also used averaging (arithmetic mean) in combining the individual expert's ratings, their paper does not specify how many experts were involved or whether they all had consistent evaluations. On the other hand, Moeinaddini et al. (2010) and Nekhay et al. (2009) combined the individual expert's ratings by calculating the geometric mean. Although either the arithmetic or geometric mean can be used in aggregating individual expert's ratings, Forman and Peniwati (1998) argue that the geometric mean is more consistent when averaging both judgments and priorities in AHP. In applying either the arithmetic or geometric mean as an aggregation method for individual preferences, the inherent assumption is that the individual evaluations of the experts would be treated equally in terms of weight or influence, irrespective of their levels of consistency, as long as they are within the acceptable limit.

Other MCDA-related observations from recent papers are connected with the application of AHP weights, its hierarchical set-up, and the criteria standardization process vis-à-vis the overall implementation of the integrated AHP and GIS methods. For example, Sener et al. (2010) categorized the criteria in landfill site suitability analysis into main criteria (stage 1), criteria (stage 2), and sub-criteria (stage 3), which were subsequently given their respective AHP weights. However, some inconsistencies have been observed, i.e., some criteria and sub-criteria have AHP weights that sum to 1, while others do not (three sets of sub-criteria at stage 3 and one set of criteria at stage 2). While two of the three sets of sub-criteria with AHP weights that do not sum to 1 may possibly be typographical errors, it is more difficult to explain the case of the third set of sub-criteria at stage 3 and the set of criteria at stage 2 (see Sener et al. 2010, p. 2041, Table 2).

Ying et al. (2007) stated that the matrices of layers B and C, and C and D were generated by the same method as for layers A and B, where the AHP weights of the factors at layer B had a total of 1. As a result, all the sub-factors at layers C and D, irrespective of the factors to which they belong in the higher hierarchy (i.e., layer B for layer C, and layer C for layer D), also had a total of 1. This method of AHP weights calculations is quite different from the method used by other researchers (e.g., Hossain et al. 2009; Khoi and Murayama 2010; Sener et al. 2010), where the weights of the given sub-factors (with a total weight of 1) under a particular factor are calculated with respect to that factor. If the calculations had been done this way, the weights of the sub-factors at layer C should have been computed with respect to their corresponding factors at layer B (and likewise for the sub-factors at layer D with respect to their corresponding factors at layer C). This method could have given the sub-factors at layer C that are under a certain factor at layer B a total weight of 1. However, this was not the case. On the other hand, in their landslide susceptibility analysis, Kamp et al. (2008) extensively discussed how much the area was affected by landslides during the 2005 earthquake vis-à-vis each criterion under consideration. However, the criteria standardization method and the overall aggregation procedures, which are vital in this kind of research, were not fully explained or disclosed.

In summary, the findings of this review show that researchers have still not reached a consensus on some issues related to the implementation of AHP as a weighting method for GIS-based MCDA. The contentious areas include: (1) the method for capturing experts' opinions using the pair-wise comparison method, (2) the method for aggregating individual expert ratings (in cases where consensus ratings are not used), and (3) the method for standardizing the criteria or factors involved in the analysis. These issues are crucial to the use of AHP for GIS-based MCDA, and are discussed further below.

Nevertheless, since it is both precise and easy to use, the AHP method has received considerable attention because it also places greater emphasis on the structure of the preferences of the decision makers (Kamp et al. 2008; Schmoldt et al. 2001). Its strength lies in its ability to capture human preferences in a hierarchical framework that focuses on the selection, definition, and measurement of criteria for a single goal or outcome (Itami et al. 2000). However, the large number of pair-wise comparisons needed in the hierarchy has been a major drawback of the AHP (Carmone et al. 1997; Islam and Abdulla 2006). As the size of the hierarchy increases, the number of pair-wise comparisons increases rapidly. Because of this, the respondents in the pair-wise comparison process are likely to suffer from information overload, even under the best circumstances (Carmone et al. 1997). This has been a major criticism of AHP. Furthermore, pair-wise comparisons are subjective, and the quality of the results is highly dependent on the expert's judgment

(Kamp et al. 2008). Therefore, a method that improves on the process of capturing experts' preferences is needed.

This review shows two major ways that are used to gather experts' opinions: consensus and individual. According to Kim and Min (2004), a consensus derived from a series of discussions is the best way. Islam and Abdulla (2006) also proposed the use of a nominal group to identify the insignificant criteria that might have been included when solving large-scale enterprise multi-criteria decision-making problems that involve large numbers of criteria. Consequently, irrelevant criteria could be dropped from the subsequent analysis, and this exclusion would not significantly affect the final decision. Eastman et al. (1995) also suggest focus-group discussions as a way to derive consensus when evaluating different criteria/factors. Hossain et al. (2009) presented the use of two analytical procedures to help reduce some of the subjectivity, to verify the weights generated, and to reach a consensus on weights. These are the use of questionnaires, and a group discussion to derive the final weight consensus. According to Robbins (1994), the size of a decision-making group can be from a minimum of 5 to a maximum of about 50. However, it must be remembered that when using consensus ratings, participants in the focus group may not necessarily have the same level of expertise and influence. Aside from the fact that this procedure is time-consuming, participants in positions of authority may also influence the other participants, thus resulting in a biased outcome (Kim and Min 2004).

In cases where consensus ratings are not met and individual expert's ratings will be used in the analysis, a method to combine such ratings is needed. This is to ensure that the important input of the experts involved will not be lost. The case reported by Hossain et al. (2009) is a typical example of the experts' opinions being substantially lost when they were discarded despite having CRs that were within the acceptable limit favoring the first author's own evaluation. While lower inconsistency is preferred, this does not mean that it is more accurate. Vaidya and Kumar (2006) considered the use of the geometric mean of the element scores from a questionnaire to arrive at the final solution. Forman and Peniwati (1998) introduced the weighted geometric mean and weighted arithmetic mean as possible methods of combining individual experts' judgments and priorities in AHP. These are in addition to the simple arithmetic and geometric mean methods. Furthermore, Kim and Min (2004) argue that the CR can be used to assess the level of expertise of the respondents. Therefore it can also be used as a basis to combine individual expert's ratings. As previously mentioned, however, lower inconsistency does not necessarily imply high accuracy. In other words, although the criteria or factors in a given problem are ranked logically and consistently, this does not guarantee high accuracy in terms of how these ranked criteria or factors contribute to the attainment of the goal or specific objective. This needs to be considered when using the CR as a basis for combining experts' ratings.

Criteria standardization is equally important. Eastman et al. (1995) argue that another primary issue, aside from the weights development for the criteria, is how criterion scores are standardized. Every method can yield a different result; hence, the selection of the method to be used must be taken seriously. Local experts can give valuable support, and actual field observations may help improve existing knowledge of the area being investigated. The method used in a particular analysis should not be kept hidden (e.g., Kamp et al. 2008); instead a reasonable amount of detail should be given to allow other researchers and analysts to evaluate or verify the analysis.

11.5 Concluding Remarks and Future Prospects

Since its introduction, AHP has been an important tool for decision makers and researchers. It offers a flexible, step-by-step and transparent way of analyzing complex problems in a MCDA environment based on experts' preferences, knowledge, and judgments. In fact, the resolution of complex problems in a multi-criteria environment has been the primary use of AHP. In a real setting, slope, elevation, rainfall, accessibility, environmental impact, etc., have different dimensions, making a simple MCDA problem more complex. In a typical MCDA situation, where multiple criteria and different fields are involved, there is a need to consider multiple stakeholders and wide-ranging expertise. Owing to its ability to readily incorporate multiple judgments, AHP and its combination with other tools such as GIS offer a solution to multi-dimensional complex problems. Indeed, its application as a decision support tool or weighting method for GIS-based MCDA marks its potential usefulness to wider applications in the vast field of geospatial analysis.

However, the problems mentioned in the previous section relating to the use of AHP for GIS-based MCDA need to be overcome when considering the further development and future prospects and implementations of this integrated method. The method of capturing and aggregating experts' opinions is one that really needs careful attention. It is not enough to just report that "experts were consulted," since in most cases different experts have different preferences, expertise, ideas, or opinions. There is a need to give an overview of the method used in dealing with this aspect of AHP, i.e., whether dealt with through a consensus or by individual evaluations. Moreover, since the strength of AHP lies in its ability to capture human preferences, its results vary depending on the subjective knowledge of the experts. Therefore it is important to take into consideration the qualifications of experts to be consulted and involved in the process. This is true whether AHP is used as a standalone method or in combination with other techniques such as GIS-based MCDA. While there are several procedures that may be used to standardize the different factors in a typical GIS-based MCDA problem, the method employed should at least be reported. This is important, since it can give a clearer picture of how a certain objective can be attained by the factors (and constraints) involved, and how each factor is going to react once the AHP weight is applied.

Nevertheless, AHP applications will undoubtedly expand in both non-spatial and spatial-based decision-making environments. This is owing to the effectiveness of AHP in evaluating problems involving multiple and diverse criteria, the measurement of trade-offs (Banai 1989), its simplicity and robustness (Vargas 1990), its precision, ease of use, and availability as a built-in tool in IDRISI GIS software

(Eastman 2006, 2009) and as an extension tool in ArcGIS software (Marinoni 2004), and in Expert Choice software (http://www.ExpertChoice.com). To take advantage of the versatility of AHP, current research efforts are also focused on how AHP can be combined with other techniques. This review has shown some of the integrations of AHP with other tools/methods that have been implemented on a GIS-based MCDA platform, like its integration with GIS, WLC, and fuzzy logic. It is expected that in the vast field of geospatial analysis, more tools/methods, in combination with AHP, will be explored in the near future.

References

Banai R (1989) A new method for site suitability analysis: the analytic hierarchy process. Environ Manag 13:685–693

Banai R (1993) Fuzziness in geographical information systems: contribution from the analytic hierarchy process. Int J Geogr Inform Sci 7:315–329

Carmone FJ Jr, Kara A, Zanakis SH (1997) A Monte Carlo investigation of incomplete pair-wise comparison matrices in AHP. Eur J Oper Res 102:538–553

Carver SJ (1991) Integrating multi-criteria evaluation with geographical information systems. Int J Geogr Inform Syst 5:321–339

Chakhar S, Martel JM (2003) Enhancing geographical information systems capabilities with multi-criteria evaluation functions. J Geogr Inform Decis Anal 7:47–71

Church RL, Loban SR, Lombard K (1992) An interface for exploring spatial alternatives for a corridor location problem. Comput Geosci 8:1095–1105

Cowen D (1988) GIS versus CAD versus DBMS: what are the differences. Photogramm Eng Rem Sens 54:1551–1555

Coyle G (2004) The analytic hierarchy process (AHP). Practical strategy. Open Access Material. Retrieved on January 11, 2011 from http://www.booksites.net/download/coyle/student_files/AHP_Technique.pdf

Diamond JT, Wright JR (1988) Design of an integrated spatial information system for multi-objective land-use planning. Environ Plann 15:205–214

Eastman JR (1999) IDRISI 32: user's guide. Clark University, Worcester

Eastman JR (2006) IDRISI Andes tutorial. Clark University, Worcester

Eastman JR (2009) IDRISI Taiga: guide to GIS and image processing. Clark University, Worcester

Eastman JR, Jin W, Kyem PAK, Toledano J (1992) Participatory procedures for multi-criteria evaluation in GIS. In: Proceedings, Chinese professionals in GIS'92, NCGIA, SUNY Buffalo, Buffalo, pp 281–288

Eastman JR, Jin W, Kyem PAK, Toledano J (1995) Raster procedures for multi-criteria/multi-objective decisions. Photogramm Eng Rem Sens 61:539–547

Eastman JR, Kyem PAK, Toledano J, Jin W (1993) GIS and decision making. UNITAR, Geneva

Estoque RC, Murayama Y (2010) Suitability analysis for beekeeping sites in La Union, Philippines, using GIS and multi-criteria evaluation techniques. Res J Appl Sci 5:242–253

Feick RD, Hall BG (2004) A method for examining the spatial dimension of multicriteria weight sensitivity. Int J Geogr Inform Sci 18:815–840

Fernandez DS, Lutz MA (2010) Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Eng Geol 111:90–98

Forman E, Peniwati K (1998) Aggregating individual judgments and priorities with the analytic hierarchy process. Eur J Oper Res 108:165–169

Forman EH, Gass SI (2001) The analytic hierarchy process: an exposition. Oper Res 49:469–486

Geldermann J, Rentz O (2007) Multi-criteria decision support for integrated technique assessment. In: Kropp JP, Scheffran J (eds) Advanced methods for decision making and risk management in sustainability science. Nova, New York, pp 257–273
Guiqin W, Li Q, Guoxue L, Lijun C (2009) Landfill site selection using spatial information technologies and AHP: a case study in Beijing, China. J Environ Manag 90:2414–2421
Heywood I, Oliver J, Tomlinson S (1995) Building an exploratory multi-criteria modeling environment for spatial decision support. In: Fisher P (ed) Innovations in GIS 2. Taylor & Francis, Bristol
Hossain MS, Chowdhury SR, Das NG, Sharifuzzaman SM, Sultana A (2009) Integration of GIS and multicriteria decision analysis for urban aquaculture development in Bangladesh. Landsc Urban Plann 90:119–133
Hutchinson CF, Toledano J (1993) Guidelines for demonstrating geographical information systems based on participatory development. Int J Geogr Inform Syst 7:453–461
Islam R, Abdulla NA (2006) Management decision-making by the analytic hierarchy process: a proposed modification for large-scale problems. J Int Bus Entrepren Dev 3:18–40
Itami RM, MacLaren GS, Hirst KM (2000) Integrating the analytical hierarchy process with GIS to capture expert knowledge for land capability assessment. In: Proceedings of the 4th conference on integrating GIS and environmental modeling, Banff
Jankowski P (1995) Integrating geographical information systems and multiple criteria decision making methods. Int J Geogr Inform Syst 9:251–273
Jankowski P, Richard L (1994) Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. Environ Plann Plann Des 21:323–340
Janssen R, Rietveld P (1990) Multicriteria analysis and geographical information systems: an application to agricultural land use in The Netherlands. In: Scholten HJ, Stillwell JCH (eds) Geographical information systems for urban and regional planning. Kluwer Academic, Dordrecht
Johnson CR (1980) Constructive critique of a hierarchical prioritization scheme employing paired comparisons. In: Proceedings of the international conference of cybernetics and society of the IEE (Institute of Electrical Engineers), Cambridge
Kamp U, Growley BJ, Khattak GA, Owen LA (2008) GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology 101:631–642
Khoi DD, Murayama Y (2010) Delineation of suitable cropland areas using a GIS based multi-criteria evaluation approach in the Tam Dao National Park Region, Vietnam. Sustainability 2:2024–2043
Kim SC, Min KJ (2004) Determining multi-criteria priorities in the planning of electric power generation: the development of an analytic hierarchy process for using the opinions of experts. Int J Manag 21:186–193
Laaribi A, Chevallier JJ, Martel JM (1996) A spatial decision aid: a multicriterion evaluation approach. Comput Environ Urban Syst 20:351–366
Malczewski J (1999) GIS and multi-criteria decision analysis. Wiley, New York
Malczewski J (2004) GIS-based land-use suitability analysis: a critical overview. Prog Plan 62:3–65
Malczewski J (2006) GIS-based multicriteria decision analysis: a survey of the literature. Int J Geogr Inform Sci 20:703–726
Marinoni O (2004) Implementation of the analytical hierarchy process with VBA in ArcGIS. Comput Geosci 30:637–646
Mendoza GA (1997) Introduction to analytic hierarchy process: theory and applications to natural resources management. In: American Society of Photogrammetry and Remote Sensing, 1997 ACSM/ASPRS/RT convention, vol 4, Resource Technology, Bethesda, pp 130–139
Mendoza GA, Martins H (2006) Multi-criteria decision analysis in natural resource management: a critical review of methods and new modeling paradigms. For Ecol Manag 230:1–22
Moeinaddini M, Khorasani N, Danehkar A, Darvishsefat AA, Zienalyan M (2010) Siting MSW landfill using weighted linear combination and analytical hierarchy process (AHP) methodology in GIS environment (case study: Karaj). Waste Manag 30:912–920

Nekhay O, Arriaza M, Guzmán-Álvarez JR (2009) Spatial analysis of the suitability of olive plantations for wildlife habitat restoration. Comput Electron Agr 65:49–64

Nyerges TL, Jankowski P (2010) Regional and urban GIS: a decision support approach. Guilford Press, New York

Pereira JMC, Duckstein L (1993) A multiple criteria decision-making approach to GIS-based land suitability evaluation. Int J Geogr Inform Syst 7:407–424

Radiarta IN, Saitoh SI, Miyazono A (2008) GIS-based multi-criteria evaluation models for identifying suitable sites for Japanese scallop (*Mizuhopecten yessoensis*) aquaculture in Funka Bay, southwestern Hokkaido, Japan. Aquaculture 284:127–135

Rao M, Sastry SVC, Yadar PD, Kharod K, Pathan SK, Dhinwa PS, Majumdar KL, Kumar DS, Patkar VN, Phatak VK (1991) A Weighted index model for urban suitability assessment: a GIS approach. Bombay Metropolitan Regional Development Authority, Bombay

Robbins SP (1994) Management. Prentice Hall, New Jersey

Saaty TL (2001) Fundamentals of the analytic hierarchy process. In: Schmoldt DL, Kangas J, Mendoza GA, Pesonen M (eds) The analytic hierarchy process in natural resources and environmental decision making. Kluwer Academic, Dordrecht

Saaty TL (1980) The analytic hierarchy process. McGraw Hill, New York

Saaty TL (1988) Multi-criteria decision making: the analytical hierarchy process. RWS Publications, Pittsburgh

Saaty TL (1994a) How to make a decision: the analytic hierarchy process. Interfaces 24:19–43

Saaty TL (1994b) Fundamentals of decision making. RWS Publications, Pittsburgh

Saaty TL (2008a) Decision making for leaders: the analytic hierarchy process for decisions in a complex world. RWS Publications, Pittsburgh

Saaty TL (2008b) Decision making with the analytic hierarchy process. Int J Services Sci 1:83–98

Samani KM, Hosseiny SA, Lotfalian M, Najafi M (2010) Planning road network in mountain forests using GIS and Analytic Hierarchical Process (AHP). Caspian J Environ Sci 8:151–162

Schmoldt DL, Kangas J, Mendoza GA, Pesonen M (eds) (2001) The analytic hierarchy process in natural resources and environmental decision making. Kluwer Academic, Dordrecht

Sener S, Sener E, Nas B, Karagüzel R (2010) Combining AHP with GIS for landfill site selection: a case study in the Lake Beysehir catchment area (Konya, Turkey). Waste Manag 30:2037–2046

Siddiqui MZ, Everett JW, Vieux BE (1996) Landfill siting using geographic information systems: a demonstration. J Environ Eng 122:515–523

Suárez-Vega R, Santos-Peñate DR, Dorta-González P, Rodríguez-Díaz M (2011) A multi-criteria GIS based procedure to solve a network competitive location problem. Appl Geogr 31:282–291

Sugumaran R, Bakker B (2007) GIS-based site suitability decision support system for planning confined animal feeding operations in Iowa. In: Hilton BN (ed) Emerging spatial information systems and applications. Idea Group Publishing, Hershey, pp 219–238

Teknomo K (2006) Analytic hierarchy process (AHP) tutorial. Retrieved on January 11, 2011 from http://people.revoledu.com/kardi/tutorial/ahp/

Thapa RB, Murayama Y (2008) Land evaluation for peri-urban agriculture using analytical hierarchical process and geographic information system techniques: a case study of Hanoi. Land Use Policy 25:225–239

Thill JC (1999) Multicriteria decision-making and Analysis: a geographic information sciences approach. Ashgate, New York

Triantaphyllou E (2000) Multi-criteria decision making: a comparative study. Kluwer Academic, Dordrecht

Tudes S, Yigiter ND (2010) Preparation of land use planning model using GIS based on AHP: case study Adana-Turkey. Bull Eng Geol Environ 69:235–245

Vaidya OS, Kumar S (2006) Analytic hierarchy process: an overview of applications. Eur J Oper Res 169:1–29

Vargas LG (1990) An overview of the analytic hierarchy process and its applications. Eur J Oper Res 48:2–8

Vizarri M (2011) Spatial modelling of potential landscape quality. Appl Geogr 31:108–118

Voogd H (1983) Multicriteria evaluation for urban and regional planning. Pion, London

Wu CH, Chen SC (2009) Determining landslide susceptibility in Central Taiwan from rainfall and six site factors using the analytical hierarchy process method. Geomorphology 112:190–204

Xiang XN, Whitley D (1994) Weighting land suitability factors by the PLUS method. Environ Plann Plann Des 21:273–304

Yalcin A (2008) GIS-based landslide susceptibility mapping using an analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. Catena 72:1–12

Yang J, Shi P (2002) Applying an analytic hierarchy process in a firm's overall performance evaluation: a case study in China. Int J Bus 7:29–46

Ying X, Guang-Ming Z, Gui-Qiu C, Lin T, Ke-Lin W, Dao-You H (2007) Combining AHP with GIS in a synthetic evaluation of eco-environment quality: a case study of Hunan Province, China. Ecol Model 209:97–109

Zahedi F (1986) The analytic hierarchy process: a survey of the method and its applications. Interfaces 16:96–108

Chapter 12
GIS Network Model in Geospatial Analysis

Ko Ko Lwin and Yuji Murayama

12.1 Introduction

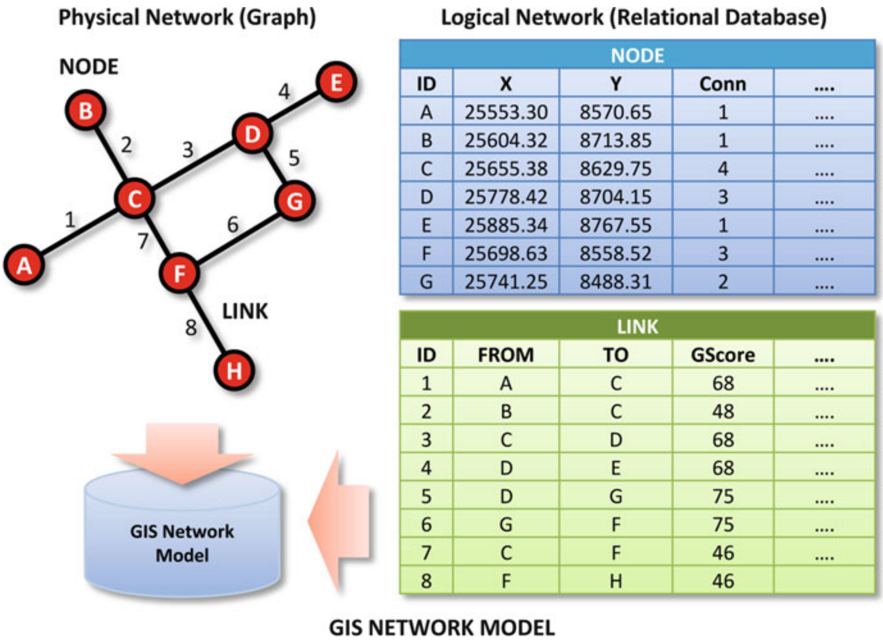
Geographic information systems (GIS) provide both theory and methods that have the potential to facilitate the development of spatial analytical functions and various GIS data models. There are several network models in GIS, such as river networks, utility networks and transportation or road networks. Among these, GIS road network data models are important for solving problems in urban areas such as transportation planning, retail market analysis, accessibility measurements, service allocation and more. Understanding the road network patterns in urban areas is important for human mobility studies, because people are living and moving along the road networks. A network data model allows us to solve daily problems such as finding the shortest path between two locations, looking for the closest facilities within a specific distance or estimating drive times. Although many network models are conceptually simple they are mathematically complex and require computational resources to model the problem.

12.2 A GIS Network Model

A network is referred to as a pure network if only its topology and connectivity are considered. If a network is characterized by its topology and flow characteristics (such as capacity constraints, path choice and link cost functions), it is referred to as

K.K. Lwin (✉) • Y. Murayama
Division of Spatial Information Science,
Graduate School of Life and Environmental Sciences,
University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kokolwin2002@yahoo.com; mura@geoenv.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_12, © Springer Japan 2012



NODE				
ID	X	Y	Conn	....
A	25553.30	8570.65	1	....
B	25604.32	8713.85	1	....
C	25655.38	8629.75	4	....
D	25778.42	8704.15	3	....
E	25885.34	8767.55	1	....
F	25698.63	8558.52	3	....
G	25741.25	8488.31	2	....

LINK				
ID	FROM	TO	GScore	....
1	A	C	68	....
2	B	C	48	....
3	C	D	68	....
4	D	E	68	....
5	D	G	75	....
6	G	F	75	....
7	C	F	46	....
8	F	H	46	

Fig. 12.1 Conceptual design of a simple GIS network model

a flow network (Fischer 2004). A transportation network is a flow network representing the movement of people, vehicles or goods (Bell and Iida 1997). Nevertheless, a common representation of a network is a set of nodes and a set of links. A network is an interconnected set of points and lines that represent possible routes from one location to another. For example, an interconnected set of lines representing city streets is a network. Therefore, the elements of a network are nodes and links. Nodes represent points of change and intersections within a network. Links are linear features made up of two or more nodes (from and to) that represent features such as roads, rivers and railroads. The attributes of nodes and links are stored in tables and assigned unique identification numbers (key fields) known as a relational database (Fig. 12.1).

When these two tables (i.e., nodes and links) are relationally linked, a basic topology can be constructed such as connectivity and a shimbel matrix. Moreover, additional attribute fields can be inserted to restrict the model, such as imposing one-way or two-way traffic, speed limitations, greenness score, etc. Many efforts have been made to create comprehensive transportation network databases to address a wide variety of transportation problems, ranging from public transit to package distribution. Initially, these efforts were undertaken within transportation network optimization packages (e.g., EMME/2, TransCAD), which created topologically sound representations. However, many of these representations were geographically inaccurate and had limited visual and geocoding capabilities. Using a network data model for the purposes of cartography, geocoding and routing requires further developments (Rodrigue et al. 2009).

In fact, a GIS network data model is based on graph theory, which is a mathematical expression used to represent aspects of the environment such as rivers, roads, subways, utilities and so on. However, a GIS network model is more than a graph; it not only represents an abstract idea of the environment, but also stores, retrieves, modifies, analyzes and displays the real-world structures in graphical form. A GIS network model includes topology, which is the arrangement of the nodes and links in the network, and the information about "which links are connected to which nodes and to other links". The ArcGIS Network Analyst software incorporates an advanced connectivity model which accurately represents real-world multi-modal networks. The integration of information about other attributes of the GIS network model allows us to analyze more complex real-world solutions such as incorporating elevation information for each node, which allows us to retrieve information about flow direction, bridges, tunnels, and overpasses and underpasses on highways.

A network can have a set of weights associated with it. A weight can be used to represent the cost of traversing an element in the logical network. For example, in a network of water pipes, a certain amount of pressure is lost when traveling the length of a transmission, mainly due to surface friction within the pipe. In transportation planning, the weighting factor could be speed limitations, traffic volumes, etc. Network weights apply to all elements in the network. The weighting values for each network element are derived from the attributes of the corresponding feature. In Chap. 16, we discuss the modeling of urban green space walkability based on a greenness score, which was derived from the normalized different vegetation index (NDVI) within a 10 m buffer of the roads. In that study, the greenness score was used as a weighting factor to identify the shortest or greenest route (i.e., the most highly vegetated route).

Solving network data problems involves complex mathematical calculations and requires high-performance computational power. Such network problems are conceptually simple, but are mathematically complex and challenging (Fischer 2004). Perhaps the most complicated model is a transportation network model in terms of complications in the attributes themselves, such as driving/speed/turning restrictions, traffic volume, etc., and involving multi-layer networks such as a railway network, a subway, or a bus network, etc. The ArcGIS software constructs the network data set by having points and lines in two separate layers (i.e., a layer-based approach) and creating additional layers for analysis such as a route analysis layer, a closest facility analysis layer, and a service analysis and OD cost matrix analysis layer (origins and destinations). Network analysis layers are composite layers in ArcMap that store the inputs, parameters and results of a network analysis. A network analysis layer acts as an in-memory workspace for each type of input as well as the result, all of which are stored as in-memory feature classes. The analysis parameters are stored as properties of the analysis layer.

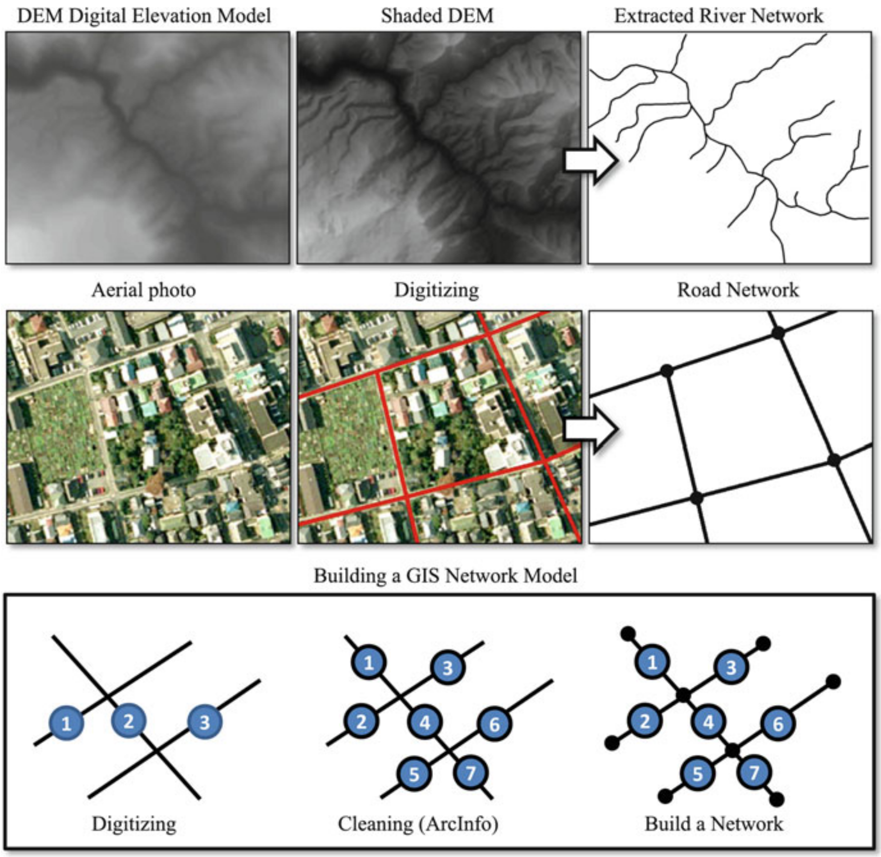


Fig. 12.2 Sources of data for the construction and building of a GIS network model

12.3 Building a GIS Network Model

A GIS network can be constructed from various sources, such as a river network, which can be extracted from a contour map or digital elevation model (DEM), or a road network, which can be constructed from aerial photos or high-resolution satellite images (Fig. 12.2). Road outline and center-line data can be purchased from GIS data vendors. It is possible to represent one-way streets in the line-theme feature based on the digitized directions in a road network. Alternatively, flow direction can be defined by the elevations of the nodes. Moreover, digitizing is based on the target application and analysis, for example, a walkability study with green space may require both sides of the main road to be digitized, since the green space may be present on each side of the road (Fig. 12.3). Multi-modal network data sets, while being more complex connectivity scenarios, are also possible, as in multi-modal

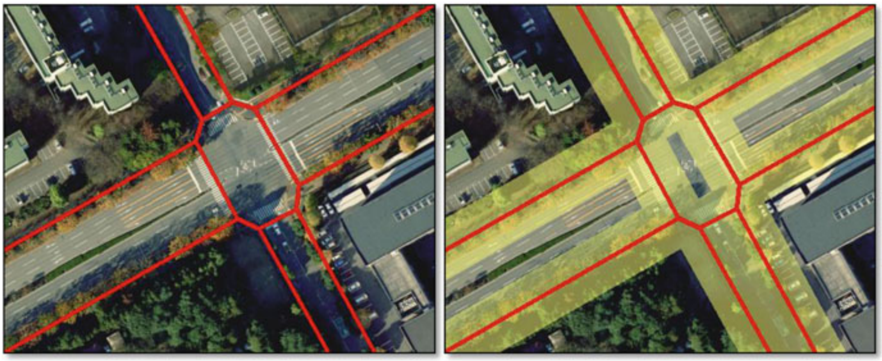


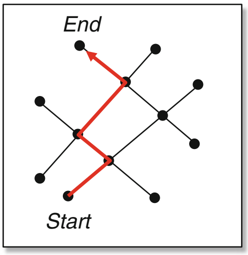
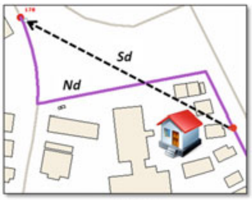
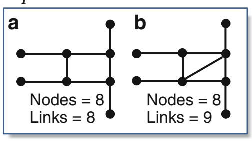
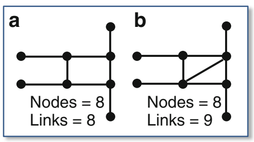
Fig. 12.3 Road digitizing for greenness score calculation within a 10 m buffered road

transportation networks. A network data set also possesses rich network attributes that helps model impedances, restrictions and hierarchy for the network.

12.4 GIS Network Model in Geospatial Analysis

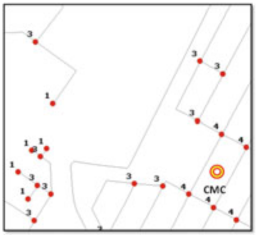
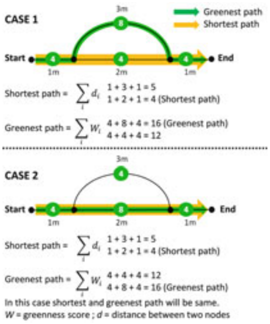
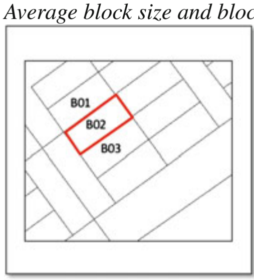
GIS network models are widely used in transportation planning, market analysis and accessibility studies. Accessibility is a measure of the spatial distribution of activities around a point, adjusted for the ability and desire of people to overcome spatial separation (Handy and Niermeier 1997). Several studies describe accessibility, review various accessibility measures, provide case studies, and present novel methods (Levinson and Krizek 2005; Bhat et al. 2002; Handy and Niermeier 1997; Pirie 1979). In short, accessibility describes travel between a source and a target in terms of factors of ease or difficulty. For example, having retail stores close to where people live, and providing connected streets, increases the likelihood that people will incorporate walking into their daily routines (Frank and Engelke 2005; Moudon et al. 2007). Furthermore, spatial syntax has been proposed as a new computational language to describe the patterns of modern cities (Hillier and Hanson 1984; Hillier 1996) based on road networks. Typical applications of spatial syntax include pedestrian modeling, crime mapping, and route-finding processes in complex built environments (Peponis et al. 1990; Hillier 1996; Jiang 1999). An axial line-based representation of an urban structure is the earliest approach to spatial syntax (Hillier and Hanson 1984). Recent developments in spatial information science have much to offer for the identification of land-use types, street connectivity, and access to services in order to determine the factors based on a GIS road network data model that might increase or decrease the probability of people being physically active according to selected spatial units of interest (Leslie et al. 2007). The properties of a network can be measured by several indices and equations. Table 12.1 shows a summary of accessibility measures based on a network model in a real-world situation.

Table 12.1 Accessibility measures based on a network data model

No.	Measurement	Description
1	*Shortest path* (network distance) 	Shortest-path analysis is a fundamental application in network analysis, and is known as network distance. Dijkstra is the most common algorithm used in finding the shortest path, which also has a lower computational complexity (Zhan and Noon 1998). Dijkstra's algorithm calculates the shortest path from a selected start node to any other node in a connected network (Dijkstra 1959). Many GIS software tools use this algorithm for shortest-path analysis
2	*Directness*  Directness = Sd/Nd	Directness is a measurement of network efficiency and complexity by rationing Sd (straight line distance) and Nd (network distance). The value is between 1 and 0, where the most efficient and simplest network has a value of 1, while a complex network has a value of 0
3	*Alpha index*  Alpha index = (Links − Nodes + 1)/(2 Nodes − 5) Alpha index for network $A = 0.09$ Alpha index for network $B = 0.18$	The alpha index is a measurement of the connectivity level. The higher the alpha index, the more connected a network is. Trees and simple networks will have a value of 0. A value of 1 indicates a completely connected network. It is very rare that an actual network will have an alpha value of 1
4	*Gamma index*  Gamma index = Links/3(Nodes − 2) Gama index for network $A = 0.44$ Gama index for network $B = 0.50$	The gamma index is a ratio of the number of links in the network to the maximum possible number of links between nodes. The maximum possible number of links is expressed as 3(Nodes − 2) because the network is abstracted as a planar graph. In a planar graph, no links intersect except at nodes (Taaffe and Gautheir 1973). This feature represents a transportation network well enough. Values for the gamma index range from 0 to 1, and are often expressed as a percentage of connectivity, e.g., a gamma index of 0.44 means that 44% of the network is connected (Dill and Portland 2003)

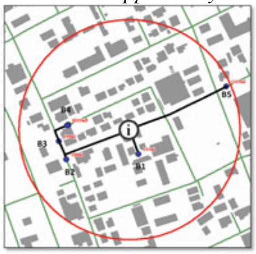
(continued)

Table 12.1 (continued)

No.	Measurement	Description
5	*Connectivity mean center* (*CMC*)  $\overline{X} = \frac{\Sigma_i W_i X_i}{\Sigma_i W_i}; \quad \overline{Y} = \frac{\Sigma_i W_i Y_i}{\Sigma_i W_i}$ W_i = number of connected streets in node *i* X_i, Y_i = node *i* coordinates $\overline{X}, \overline{Y}$ = connectivity mean center	The mean center averages the *x* and *y* coordinates of all the nodes in a specific user-defined area such as a planning zone or administrative unit. The connectivity mean center (Lwin and Murayama 2009) finds the most connected place within a specific unit based on the number of connected streets in each node as a weighted factor. It is useful for finding an optimal voting site, community center, bus stop, etc. Moreover, the degree of street connectivity can be mapped based on the values of these nodes (see Fig. 12.6)
6	*Shortest vs. greenest* 	Sometimes road attribute values can be used as weighted factors to compare shortest vs. weighted distance. For example, finding the greenest route can be accomplished based on the greenness score of each road segment (Lwin and Murayama 2011). These attribute values could also be speed limitation, number of traffic accidents, traffic volume, etc. (see Chap. 16 for more details)
7	*Average block size and block density*  Average block size = Total areas of polygons/number of polygons	Average block size (Hess et al. 1999; Reilly 2002) can be measured by converting the road line features to polygon features and calculating the average block area. A few researchers have used block density as a proxy measure for connectivity (Dill and Portland 2003). Frank et al. (2000) used the mean number of census blocks per square mile, since census blocks are typically defined as the smallest fully enclosed polygon bound by features such as roads or streams on all sides. A rule of thumb is the smaller the size, the better the accessibility

(continued)

Table 12.1 (continued)

No.	Measurement	Description
8	*Cumulative opportunity measure*  $A_i = \Sigma_j (B_j a_j)$ where A_i = accessibility measured at point i to potential activities in j a_j = opportunities in zone j B_j = a binary or threshold value	Cumulative opportunity provides a measure of the number of available facilities within a certain distance or travel time. Examples of cumulative opportunity measures are found in various articles (O'Sullivan et al. 2000; Sherman et al. 1974; Wachs and Kumagi 1973)
9	*Gravity-based measure* $L_{ij} = K(P_i P_j / d_{ij}^2)$ L_{ij} = interaction between nodes i and j P_i = magnitude of node i P_j = magnitude of node j d_{ij} = distance between two nodes K = constant (population size, total shops, total phone calls, etc.)	The gravity measure is an interaction of nodes based on their distance and some functional (weighted factors) measure of their individual accessibility. For example, weighted factors will be number of shops per point, frequency of going to these shops, etc. The gravity model is sometimes considered as a potentials model. The higher the opportunity and the shorter the distance, the greater the interaction occurring between two points, like Newton's gravity law. Down-town areas provide more opportunities for shopping and job demands, while satellite areas provide more opportunities for sight-seeing, recreation and sports activities. Hansen (1959) is generally credited with first applying the gravity approach to transportation and land-use planning

12.5 Common Network Analysis Functions in GIS

Common network analysis types include route analysis or optimal route analysis, route analysis with barriers, service area identification, and finding the closest facility and route for goods delivery (vehicle routing plan). Route analysis with barriers (Fig. 12.4) is important in mid-disaster management. The optimal route analysis can be done by either shortest distance or minimal travel time. It finds alternative routes to reach target locations by avoiding barrier points (e.g., broken roads). Many GIS packages are available for network analysis, such as the ArcGIS Network Analyst extension, SANET (http://sanet.csis.u-tokyo.ac.jp/), TranCAD, eRouteLogistics and RouteSmart. Using route analysis, we can calculate the best route from a starting point (e.g., home) to other desired points (e.g., school, hospital, etc.).

Fig. 12.4 Route analysis with a barrier in the case of a disaster scenario

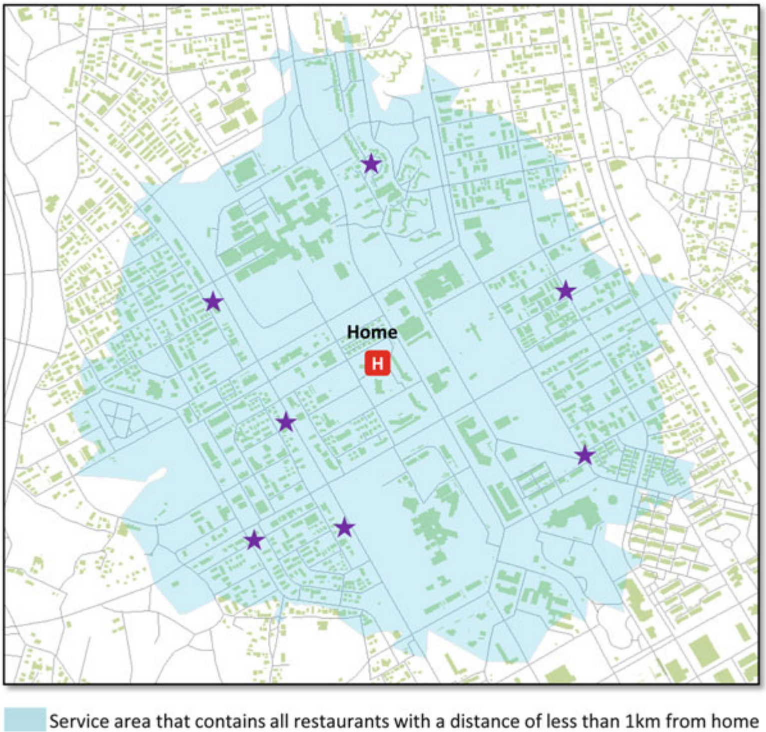


Fig. 12.5 Service area analysis

Service area identification is another common type of network analysis which identifies a region that encompasses all accessible streets within a user-defined network distance, for example finding all available restaurants within 1 km from home (Fig. 12.5).

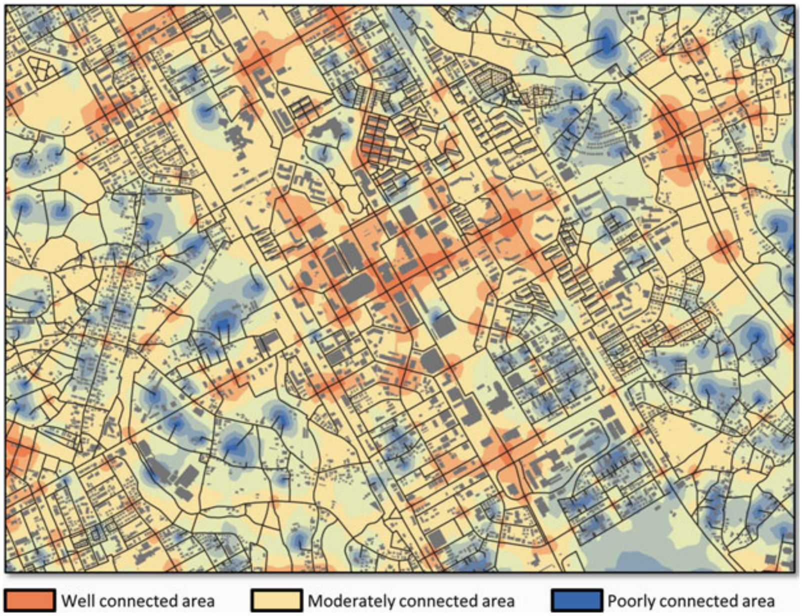


Fig. 12.6 Degree of street connectivity inside the city

Another application of a GIS road network data model is mapping the degree of street connectivity inside the city. Figure 12.6 shows the degree of street connectivity by an interpolation of connected streets for each node value. A well-connected area can be found in the city center.

Another example of a GIS network model-based application is "multi-stop trip" planning. For example, in Tsukuba City, Japan, local residents and green exercise takers can find the shortest or greenest route during their multi-stop trips (e.g., home to park, park to shopping center, shopping center to library, etc.) based on a GIS network data model through a smart phone while they are walking through the street (Fig. 12.7). By arranging the trip locations, users can calculate various travel distances and greenness scores. The increasing popularity of the Internet and user-friendly web-based GIS applications such as Google Maps/Earth and Microsoft Bing Maps platforms have made GIS an integral part of life today for finding the closest facilities, driving routes, and so on.

12.6 Conclusion

A network is a system of interconnected linear features used to solve problems of transportation, public facility management, accessibility studies and other human mobility studies. Network data sets are made of network elements. The geometry of

Fig. 12.7 Multi-stop trip planning based on a GIS network data model for mobile GIS users (http://land.geo.tsukuba.ac.jp/ecowalker/ecowalker_eng.aspx)

the source features is used to establish connectivity. In addition, network elements have attributes that control navigation within the network. Accurate GIS network analysis requires good preparation of the network data and sufficient attribute information. A network model can include a multi-layer model such as a railway system, a subway system, and a bus system to solve problems using multiple modes of transportation in an urban environment. Many commercial GIS data models composed with layers such as points (nodes) and lines (links) are using the layer-based approach. These nodes and links can also be represented as object classes in the object-oriented network data model, which is still in the design phase. The development of 3D network models and concepts in GIS (Lee 2005; Zhu et al. 2006) can solve these complex multi-layer network problems. Moreover, the combination of Internet technology and GIS gives opportunities to perform interactive network analysis in order to make spatial decisions in a timely manner for local residents and city planners (see example in Chap. 16).

References

Bell M, Iida Y (1997) Transportation network analysis. Wiley, Chichester

Bhat C, Handy S, Kockelman K, Mahmassani H, Gopal A, Srour I, Weston L (2002) Development of an urban accessibility index: formulations, aggregation, and application. Center for Transportation Research, The University of Texas at Austin. Report no. FHWA/TX-02-4938-4

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numerische Mathematik 1:269–271

Dill J, Portland OR (2003) Measuring network connectivity for bicycling and walking. In: Joint congress of ACSP-AESOP, Leuven

Fischer M (2004) GIS and network analysis. In: Hensher D, Button K, Haynes K, Stopher P (eds) Handbook of transportation geography and spatial systems. Elsevier, Amsterdam, pp 391–408

Frank LD, Engelke P (2005) Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution. Int Reg Sci Rev 28:193–216

Frank LD, Stone B Jr, Bachman W (2000) Linking land use with household vehicle emissions in the central Puget Sound: methodological framework and findings. Transp Res Part D 5: 173–196

Handy S, Niermeier DA (1997) Measuring accessibility: an exploration of issues and alternatives. Environ Plann A 29:1175–1194

Hansen WG (1959) How accessibility shapes land use. J Am Plann Inst 25:73–76

Hess PM, Moudon AV, Snyder MC, Stanilov K (1999) Site design and pedestrian travel. Transp Res Rec 1674:9–19

Hillier B (1996) Space is the machine: a configurational theory of architecture. Cambridge University Press, Cambridge

Hillier B, Hanson J (1984) The social logic of space. Cambridge University Press, Cambridge

Jiang B (1999) SimPed: simulating pedestrian crowds in a virtual environment. J Geogr Inf Decis Anal 3:21–30

Lee J (2005) 3D GIS in support of disaster management in urban areas. Retrieved from http://www.directionsmag.com/article.php?article_id=2049

Leslie E, Coffee N, Frank L, Owen N, Bauman A, Hugo G (2007) Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health Place 13:111–22

Levinson DM, Krizek KJ (2005) Access to destinations. Elsevier, Kidlington

Lwin KK, Murayama Y (2009) Interactive online micro-spatial population analysis based on GIS estimated building population. In: URISA's 2009 annual conference & exposition, Anaheim (student paper award)

Lwin KK, Murayama Y (2011) Modelling of urban green space walkability: eco-friendly walk score calculator. Comput Environ Urban Syst 35:408–420

Moudon AV, Lee C, Cheadle AD, Garvin C, Johnson DB, Schmid TL, Weathers RD (2007) Attributes of environments supporting walking. Am J Health Promot 21:448–459

O'Sullivan D, Morrison A, Shearer J (2000) Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. Int J Geogr Inf Sci 14:85–104

Peponis J, Zimring C, Choi YK (1990) Finding the building in wayfinding. Environ Behav 22:555–590

Pirie GH (1979) Measuring accessibility: a review and proposal. Environ Plann A 11:299–312

Reilly MK (2002) The influence of urban form and land use on mode choice: evidence from the 1996 bay area travel survey. Presented at the annual meeting of the Transportation Research Board, Washington

Rodrigue J-P, Comtois C, Slack B (2009) The geography of transport systems, 2nd edn. Routledge, New York

Sherman L, Barbara B, Kondo W (1974) Method for evaluating metropolitan accessibility. Transp Res Rec 499:70–82

Taaffe EJ, Gautheir H Jr (1973) Geography of transportation. Prentice-Hall, Englewood Cliffs

Wachs M, Kumagi TG (1973) Physical accessibility as a social indicator. Socio Econ Plann Sci 7:437–456

Zhan FB, Noon CE (1998) Shortest path algorithms: an evaluation using real road networks. Transp Sci 32:65–73

Zhu Q, Li Y, Tor YK (2006) 3D dynamic emergency routing. GIM-Int 20(6). http://www.gim-international.com/issues/articles/id674-D_Dynamic_Emergency_Routing.html

Part III
Applications in Geospatial Analysis

Chapter 13
Urban Growth Modeling Using the Bayesian Probability Function*

Rajesh Bahadur Thapa and Yuji Murayama

13.1 Introduction

Urban growth is recognized as physical and functional changes due to the transition of rural landscapes to urban forms. The time–space relationship plays an important role in understanding the dynamic process of urban growth. This dynamic process consists of a complex nonlinear interaction between several components, i.e., topography, rivers, land use, transportation, culture, population, economy, and growth policies. Many efforts have been made to improve such dynamic process representations with the utility of cellular automata (CA) coupled with fuzzy logic (Liu 2009), artificial neural networks (Li and Yeh 2002; Almeida et al. 2008), Markov chains with a modified genetic algorithm (Tang et al. 2007), weight of evidence (Soares-Filho et al. 2004), nonordinal and multi-nominal logit estimators (Landis 2001), SLEUTH (Clarke et al. 1997; Jantz et al. 2010), and others (White and Engelen 1997; Batty et al. 1997).

*This chapter has been improved from "Rajesh Bahadur Thapa and Yuji Murayama (2010), Urban growth modeling of Kathmandu metropolitan region, Nepal. *Computers, Environment and Urban Systems*, 35(1), 25–34," Copyright (2011), with permission from Elsevier.

R.B. Thapa (✉)
Earth Observation Research Center, Space Applications Mission Directorate,
Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: rajesh.thapa@jaxa.jp

Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: mura@geoenv.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_13, © Springer Japan 2012

Models based on the principles of CA have been developing rapidly. The CA approach provides a dynamic modeling environment which is well suited to modeling complex environments composed of a large number of individual elements. The land-use change and urban growth processes can be compared with many aspects of the behavior of a cellular automaton, for instance, the space of an urban area can be regarded as a combination of a number of cells, each cell taking a finite set of possible states representing the extent of its urban development with the state of each cell, and evolving in discrete time steps according to local transition rules. Therefore, CA-based urban models usually pay more attention to simulating the dynamic process of urban development and defining the factors or rules driving the development (Batty et al. 1997). Different CA models have been developed to simulate urban growth and urban land use/cover change over time. The differences among various models exist in modifying the five basic elements of CA, i.e., the spatial tessellation of cells, the states of cells, the neighborhood, the transition rules, and the time (Liu 2009). CA models have been shown to be effective platforms for simulating dynamic spatial interactions among biophysical and socio-economic factors associated with land-use and land-cover change (Jantz et al. 2010).

While new urban models have provided insights into urban dynamics, a deeper understanding of the physical and socio-economic patterns and processes associated with urbanization is still limited in developing countries in South Asia. Although emerging geospatial techniques have recently bridged the spatial data gap, there are still very few empirical case studies (Thapa and Murayama 2009). This research aims to simulate urban growth in the Kathmandu metropolitan region in Nepal using the weight of evidence technique incorporating CA. As the result of population growth and migration from rural to urban areas, urbanization has been recognized as a critical process in metropolitan areas of Nepal (Portnov et al. 2007; Haack 2009; Bhattarai and Conway 2010). The Kathmandu metropolitan region, which is the capital and major tourist gateway, has been facing rapid urbanization over the last three decades. Recently, it has had an estimated population of 2.18 million with an annual growth rate of 5.2% (Thapa and Murayama 2010). Such urbanization pressure results in rapid changes in the urban landscape pattern of the region, with more built-up areas and the loss of natural landscapes.

Kathmandu, the capital of Nepal, has a long history of development, and is typical of cities surrounded by complex mountain terrains in the Himalayan region. History has witnessed its development as a strategic center of power, politics, culture, and commerce (Thapa et al. 2008). However, along with the establishment of modern transportation infrastructures bringing easy access to the city, the agglomeration of rural settlements in the Kathmandu valley encroaching into the city began in the early 1960s. The predominantly agricultural landscape gradually changed to an urban landscape, with increasing human settlement in the 1960s and 1970s. These changes have escalated since the 1980s. The spatial diffusion of urban/built-up areas has spread outward from the city core and along the major roadways. Agricultural encroachment in rural hills and the periphery of mountains, and urbanization in the valley floor area, are identified as the most common phenomenon in the valley (Thapa and Murayama 2009; Haack 2009; Bhattarai and Conway 2010).

Several urban land-use development planning and policy initiatives for the valley have been made by the government in the past decades (Thapa et al. 2008). The latest planning document, "Long-Term Development Concept for the Kathmandu Valley" (Kathmandu Uptyakako Dirghakalin Bikas Avadharana), was released in 2002 (KVUDC 2002). This document, as a planning reference, conceptualizes scenarios to develop the Kathmandu metropolitan region by 2020. This long-term plan recommends: the promotion of the tourism-led service sector, guided urban development encouraging a compact urban form and the conservation of agricultural land, infrastructure development coordinated with land use, a new outer ring road to connect the traditional settlements in the metropolitan region, and rigorous regulation of areas defined as environmentally sensitive. All these policy recommendations will eventually affect the future spatial pattern of urbanization.

13.2 Methods

13.2.1 Study Site

The study area which has been selected to apply the urban growth model (Fig. 13.1) follows the watershed boundary, which was derived from 20 m digital elevation data. The topography rises to an elevation of 1,100–2,700 m above sea level and forms a bowl-shaped valley. As most of the areas outside the watershed boundary contain high mountains, forest, shrub land, and very little human settlement, urban expansion outside this boundary is largely restricted by these natural barriers. The valley is drained by the Bagmati river system, which is the main source of water for drinking and irrigation (Thapa and Murayama 2009). The study area covers 685 km^2, and 14% of the land is defined as urban area that includes five urban centers, Kathmandu, Lalitpur, Bhaktapur, Kirtipur, and Madhyapur Thimi. In addition, the region consists of 97 suburban and rural villages.

13.2.2 Database Preparation

In this study, we used data from various sources for modeling, calibrating, and validating urban growth in the Kathmandu metropolitan region. Three land-use maps at 30 m spatial resolution for the years 1978, 1991, and 2000 were processed. These maps were acquired from Thapa (2009), and were created using remote sensing techniques. The heterogeneous and complex landscape of urban regions, for example, suburban residential areas which form a complex mosaic of trees, lawns, roofs, concrete, and asphalt roadways, requires land-use and land-cover classification

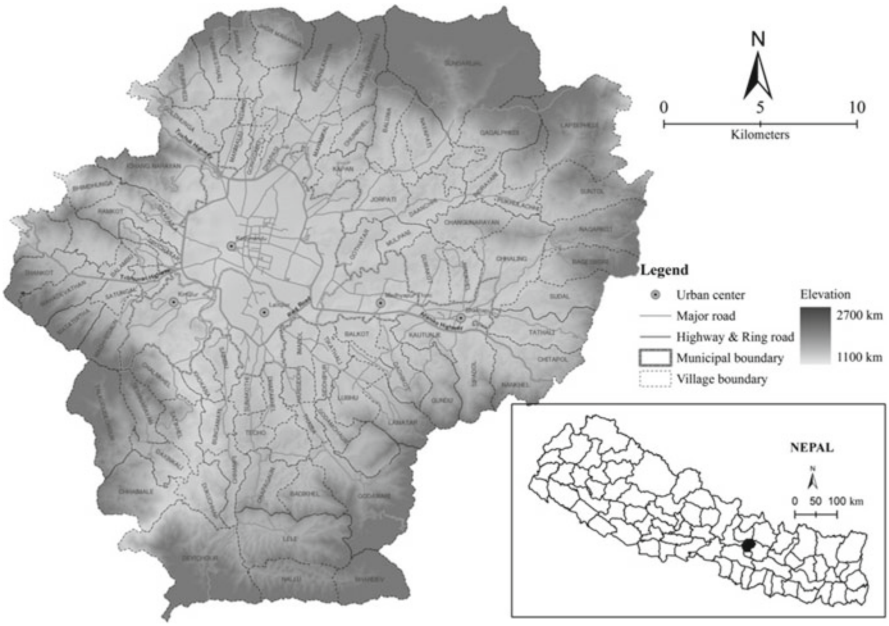


Fig. 13.1 Study area—Kathmandu valley, Nepal. *Data source*: ICIMOD/UNEP (2001)

techniques that combine more than one classification procedure in order to improve the accuracy of remote-sensing-based mapping. Therefore, a hybrid approach with a series of processing steps was developed to create multi-temporal thematic maps of the Kathmandu valley. A detailed discussion of the methods adopted to create these maps and the mapping accuracies can be found in Thapa and Murayama (2009). In this chapter, it is assumed that these reference maps are the most accurate maps available at this particular time, so they serve as the basis for measuring the accuracy of the prediction. The maps were further generalized to reduce the complexity of urban growth modeling. The land-use types, i.e., built-up areas, industrial areas, roads, airport, institutional areas, government secretariat area, and royal palace, were aggregated into a built-up category. The other land-use types, namely agricultural area, forest, shrubs, water, and open space, were kept in the same state. Protected areas were not separated from the land-cover maps, but were merged according to their nature, i.e., forest cover, open space, water, or built-up area. Two historical transition matrices were calculated from the land-use maps for the periods 1978–1991 and 1991–2000. Based on a transition matrices analysis, the water and open space areas were excluded in dynamic modeling, as they represented small land areas and mostly remained static. Therefore, for the dynamic modeling of urban growth we restricted ourselves to four broad land-cover categories, i.e., built-up areas, agriculture, forest, and shrubs. Furthermore, the land-cover transition rates for each category were computed by normalizing the row sum to be equal to 1. As the model was set to run in yearly time-steps, the transition rates were further converted to annual rates by simply dividing by the number of years, i.e., 13 and 9 for the land-cover transition periods of 1978–1991 and 1991–2000, respectively. The transition rates per year were incorporated into the modeling as fixed parameters.

Other data, i.e., elevation, urban and village boundaries, and population census, were also prepared. After creating all the required input maps in ArcGIS software, a simulation model of urban growth was designed in DINAMICA, a spatially explicit CA-based modeling software (http://www.csr.ufmg.br/dinamica/dinamica.html). The transition matrices were passed on to the DINAMICA model, which allocates the changes across the landscape based on spatial data layers representing physical and socio-economic conditions which are stored in a geographical information science (GIS) environment. As interactions between landscape elements occur in different ways, depending on local characteristics and transition rates, the DINAMICA model produces distinct spatial patterns of land-cover change.

13.2.3 Determining the Factors of Urban Growth

The selection of the best set of input variables and internal software parameters in order to produce the best fit between the empirical data and the observable reality are very important aspects of urban growth modeling. After calculating the existing land transitions, we identified a different set of factors governing each change in the four land-cover categories. In general, the urban growth that causes land-use changes is the result of a complex interaction between behavioral and structural factors associated with demand, technological capacity, and the social relations which affect demand and capacity and ultimately strain the environment (Thapa and Murayama 2010). The factors which are available for the modeling analysis do not always represent the set of variables which are necessary to produce ideal simulation results. However, there are no universal factors driving the change. Although similar driving factors have been found in several studies, the degree to which they contribute to landscape change differs (Batty et al. 1997; Almeida et al. 2008; Jantz et al. 2010). People, government plans and programs, landforms, landscape change processes, and available resources often cause differences in the importance of various factors. In the Kathmandu Valley, similarities among these factors are apparent. In fact, the people's behavior and daily interactions with the environment over time have caused observable changes in the valley's landscape (Thapa and Murayama 2010). Indeed, there is a set of factors for land-use transitions that substantially respond to landscape changes where these factors effectively guide the modeling experiment.

Several maps of biophysical features, infrastructure, and social factors have been generated on the basis of the information extracted from land-use maps and other data sources (Table 13.1). A digital elevation model (DEM) map was created based on elevation point data (ICIMOD/UNEP 2001), a slope map was derived from the DEM map, and an annual population growth-rate map was prepared based on the census data (CBS 2001). In all cases, distances were calculated using the Euclidean distance method. The distance to existing built-up areas was defined as dynamic,

Table 13.1 Definition of the land cover change driving factors

Land-change factors (biophysical, infrastructure, and social)	Year
Digital elevation model at 30 m spatial resolution	1995
Slope in degrees	1995
Distances to rivers	1978, 1991, 2000
Distances to industrial estates	1978, 1991, 2000
Distances to five urban centers (Kathmandu, Lalitpur, Kritipur, Bhaktapur, and Madhayapur Thimi)	1978, 1991, 2000
Distances to major roads and highways	1978, 1991, 2000
Distances to ring road	1978, 1991, 2000
Distances to feeder roads	1978, 1991, 2000
Distances to existing built-up surface	1978, 1991, 2000
Annual population growth rate	1991, 2000

and was up-dated at each model iteration step because of the neighborhood characteristics of land-use change attraction. All the remaining factors were static. The spatial independence of the input factor maps was checked using a set of measures, i.e., the Cramer test and the joint uncertainty information (Bonham-Carter 1994). Both tests presented a value between 0 and 1, showing the degree of association as from independent to full, respectively, between the maps compared. As a principle, correlated variables must be removed or combined into a third that will be used in the model. Among the factors selected, no significant spatial dependency was revealed where Cramer's test and the joint uncertainty information were found to be <0.5 and <0.6, respectively.

13.2.4 Transition Probability Map Calculation and Model Calibration

The development of abstraction methods capable of adequately representing complex processes with respect to quantity and location is a great challenge (Godoy and Soares-Filho 2008). The weight of evidence, entirely based on the Bayesian approach of conditional probability, is traditionally used by geologists to point out areas favorable for geological phenomena such as seismicity and mineralization (Goodacre et al. 1993; Bonham-Carter 1994). This method can combine spatial data from a variety of sources to describe and analyze interactions, provide evidence for decision making, and make predictive models (Soares-Filho et al. 2004; Almeida et al. 2008). In our spatial context, this approach detects the favorability of a certain event, for example, an event of land-cover change from agriculture to built-up surface in relation to potential evidence (proximity to urban centers, roads, water, etc.), often called the driving factors of change. Weights are estimated from the measured association between the land cover change occurrences and the values on the driving factors maps which are to be used as predictors. In this research, we employed the weight of evidence method to select the most important variables needed for the land cover change analysis and quantified their influence on each type of land cover transition event. The weights of evidence represent each variable influence on the spatial probability of a transition i⇒j, and can be calculated as follows:

$$O\{D \mid B\} = O\{D\}\frac{P\{B \mid D\}}{P\{B\bar{D}\}} \tag{13.1}$$

where $O\{D \mid B\}$ is the odds ratio of event D occurring given a spatial pattern B, $O\{D\}$ is the prior odds ratio of event D, and $P\{B \mid D\}/P\{B\bar{D}\}$ is known as the sufficiency ratio. In the weight of evidence, the natural logarithm of both sides of 13.1 is taken as

$$\log\{D \mid B\} = \log\{D\} + W^{+} \tag{13.2}$$

where W^{+} is the weight of evidence of event D occurring, which is calculated from the data. The spatial probability of a transition $i \Rightarrow j$, given a set of spatial data, can be expressed as

$$P\left(i \Rightarrow j\frac{x,y}{v}\right) = \frac{e^{\sum_k W^{+}_{kn_{i\Rightarrow j(V)x,y}}}}{1+\sum_{ij} \mathrm{e}^{\sum_k W^{+}_{kn_{i\Rightarrow j(V)x,y}}}} \tag{13.3}$$

where V is a vector of k spatial driving factors, measured at location x, y and represented by its weights $W^{+}_{1xy}, W^{+}_{2xy}, \ldots, W^{+}_{nxy}$, where n is the number of categories of each factor k (for a more detailed mathematical discussion, see Bonham-Carter 1994; Soares-Filho et al. 2004). In this way, weights of evidence are assigned for categories of each factor represented by its spatial data layers. After creating local transition probabilities, the CA simulation model is calibrated by internal parameters which concern the average size and variance of patches and patch isometry (Table 13.2). These functions enable the formation of a variety of sizes and shapes of patches of change. The patch isometry varies from 0 to 2. The patches assume a more isometric form as this number increases. The sizes of patches of change are set according to a log–normal probability distribution. Therefore it is necessary to specify the parameters of this distribution, which is represented by the mean and variances of the patch sizes to be formed (Soares-Filho et al. 2004). The mean patch size and variances were determined from the source maps, while the isometry was determined empirically. The calibration parameters were computed for three temporal years, i.e., 1978 for calibrating the predictive model of 1991 (time 1), 1991 for the model of 2000 (time 2), and 2000 for projecting the future land-cover patterns. Furthermore, the reference map of time 2 was not used during the model calibration, i.e., the land-cover map of 2000 was not used as an input map while calibrating the model of 2000.

Table 13.2 Simulation internal parameters

	1978			1991			2000		
Land cover	MPS	Var	Iso	MPS	Var	Iso	MPS	Var	Iso
Shrubs	81	21457	1	62	13849	1.1	69	15842	1.1
Forest	27	3433	1	41	8245	1.1	40	7510	1.1
Built-up	31	3755	1	7	114	1.1	6	222	1.1
Agriculture	49	5682	1	20	2573	1.1	9	438	1.1

Note: *MPS* mean patch size, *Var* variance, *Iso* isometry

13.2.5 Model Validation

The validation of a landscape dynamics model is usually carried out by comparing the predicted result with the empirical map to determine the predictive ability of the model. In this chapter, we used a three-map comparison approach for model validation, as is recommended by Pontius et al. (2008). This validation technique considers the overlay of all three maps: the reference map of time 1, the reference map of time 2, and the prediction map of time 2. This three-map comparison approach allows one to distinguish the pixels that are correct due to persistence from the pixels that are correct due to change. In this chapter, two validation maps were created for the 1991 and 2000 models. For further quantitative clarification of each model, the sources of percentage correct and percentage error were analyzed by computing observed change, predicted change, figure of merit (13.4), producer's accuracies (13.5), user's accuracies (13.6), and overall accuracies (13.7).

The figure of merit (FoM) is the ratio of the intersection of the observed change and the predicted change to the union of the observed change and the predicted change. The FoM can range from 0%, meaning no overlap between observed and predicted change, to 100%, meaning a perfect overlap between observed and predicted change. Equation 13.4 shows the mathematical notation of the FoM.

$$\mathrm{FoM} = \frac{B}{A + B + C + D} \tag{13.4}$$

where A is the area of error due to observed change predicted as persistence, B is the correct area due to observed change predicted as change, C is the area of error due to observed change predicted as the wrong gaining category, and D is the area of error due to observed persistence predicted as change.

The producer's accuracy (PA) (13.5) shows the proportion of pixels that the model predicts accurately as change, given that the reference maps indicate the observed change. The user's accuracy (UA) (13.6) shows the proportion of pixels that the model predicts accurately as change, given that the model predicts change. The overall accuracy (OA) (13.7) provides the overall agreement between the reference and predicted maps.

$$\mathrm{PA} = \frac{B}{A+B+C} \tag{13.5}$$

$$\mathrm{UA} = \frac{B}{B+C+D} \tag{13.6}$$

$$\mathrm{OA} = \frac{B+E}{A+B+C+D+E} \tag{13.7}$$

where E in 13.7 is the correct area due to observed persistence predicted as persistence.

13.3 Results

13.3.1 Land Cover Transition Analysis

The land cover transition matrix provides an important basis for analyzing the temporal and spatial changes of land cover, and examining the driving forces behind those changes in the Kathmandu metropolitan region. Figure 13.2 shows the landscape transition maps for the two time periods, i.e., 1978–1991 and 1991–2000. The maps demonstrate substantial landscape transitions during the study period. Agricultural areas gained a large amount of land at the expenses of shrubs and forest land during the period 1978–1991 (Table 13.3). Large proportions of shrub (36 km^2) and forest (26 km^2) land were transformed into agricultural land in the surrounding rural mountain areas in the region. This can be observed mostly in the northeastern and southern parts of the region.

The built area received 23 km^2 from agricultural land and 1.2 km^2 from forest. The development of a ring road around the existing urban core during the 1970s, and the extension of major and feeder roads into rural areas in the 1980s (Fig. 13.2a), accelerated the expansion of built-up areas at the expense of agricultural areas. Shrub and forest land also contributed to built-up areas, albeit at a lower rate. Depending on the location, land cover transitions between forest and shrub land were also observed.

During the period 1991–2000, a large amount of agricultural land (27 km^2) was transformed into built-up area, which increased by 4 km^2 as compared with the previous 13-year period (Table 13.4). At the same time, the transformation of shrub and forest land cover into agricultural land decreased noticeably. Some of the forest (1 km^2) and shrub (1.5 km^2) land was also changed to built-up area owing to the expansion of rural roads in the 1990s. Overall, agricultural encroachment in the periphery of rural hills and mountains mainly occurred where shrub and forest

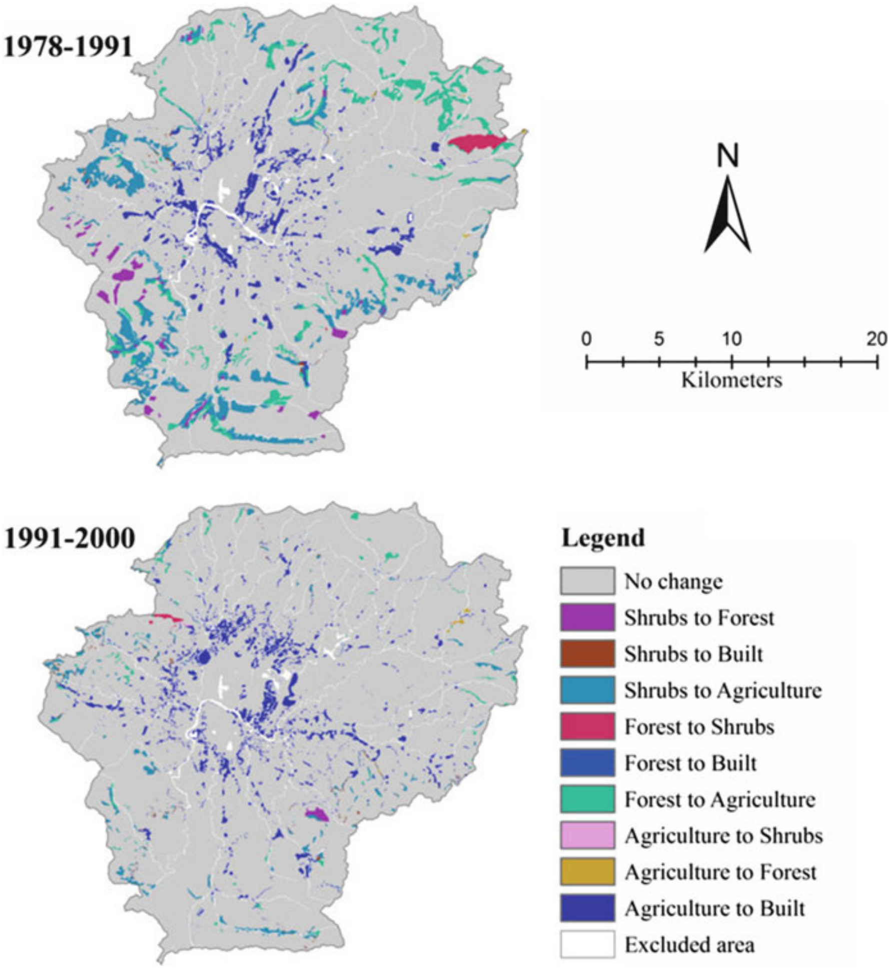


Fig. 13.2 Land cover transitions (1978–1991 and 1991–2000)

Table 13.3 Land cover transition area matrix (1978–1991)

	1991							
	Shrubs		Forest		Built-up		Agriculture	
1978	km²	Rate	km²	Rate	km²	Rate	km²	Rate
Shrubs	78.040	0.645	6.320	0.052	0.570	0.005	36.090	0.298
Forest	3.140	0.019	132.220	0.814	1.200	0.007	25.970	0.160
Built-up	0.000	0.000	0.000	0.000	32.100	1.000	0.000	0.000
Agriculture	0.000	0.000	0.420	0.001	22.920	0.065	328.520	0.934

Table 13.4 Land cover transition area matrix (1991–2000)

	2000							
	Shrubs		Forest		Built-up		Agriculture	
1991	km^2	Rate	km^2	Rate	km^2	Rate	km^2	Rate
Shrubs	71.070	0.876	0.920	0.011	0.950	0.012	8.210	0.101
Forest	0.540	0.004	132.630	0.957	1.490	0.011	3.860	0.028
Built-up	0.000	0.000	0.000	0.000	57.060	1.000	0.000	0.000
Agriculture	0.080	0.000	0.450	0.001	27.150	0.069	363.420	0.929

landscape in rural areas of the valley changed to agricultural areas, while in the valley floor, the conversion of agricultural land into built-up surfaces was identified. A minor land-use transition between forest and shrub land cover was also noticed during this period.

13.3.2 Urban Growth Model Validation Results

By varying the parameters at each model iteration, various simulation results were produced. The best results generated by the model are illustrated in Fig. 13.3c, e, which are compared with the actual maps of land cover from 1978, 1991, and 2000 (Fig. 13.3a, b, d). The similarity in the spatial patterns between the simulated and reference maps are very important. A visual comparison of the model's simulated results and the actual maps from 1978–2000 shows that the results produced by the model were a good match with the actual urban events. However, a systematic validation method requires us to quantify the degree of error in the simulation results. Figure 13.4 shows the validation results graphically, and allows the reader to access the nature of the prediction errors, which are shown in various colors. These are obtained by overlaying the reference map of time 1, the reference map of time 2, and the prediction map of time 2. The medium-pink pixels show where the model predicted change correctly. Purple pixels show where change was observed and the model predicted change, but the model predicted a transition to the wrong category, which is a type of error that can occur in multiple land-cover category models. Medium-blue pixels show where change was observed at locations where the model predicted persistence. Light-blue pixels show where persistence was observed at locations where the model predicted change. White pixels show locations where the model predicted persistence correctly.

Figure 13.5 presents a summary of the error analysis according to the logic of the legend for Fig. 13.4. Each bar is a rectangular Venn diagram where the two central segments with a different gray scale represent the intersection of the observed change and the predicted change. The second segment from the left shows the change that the model predicted correctly. The union of the segments on the left and

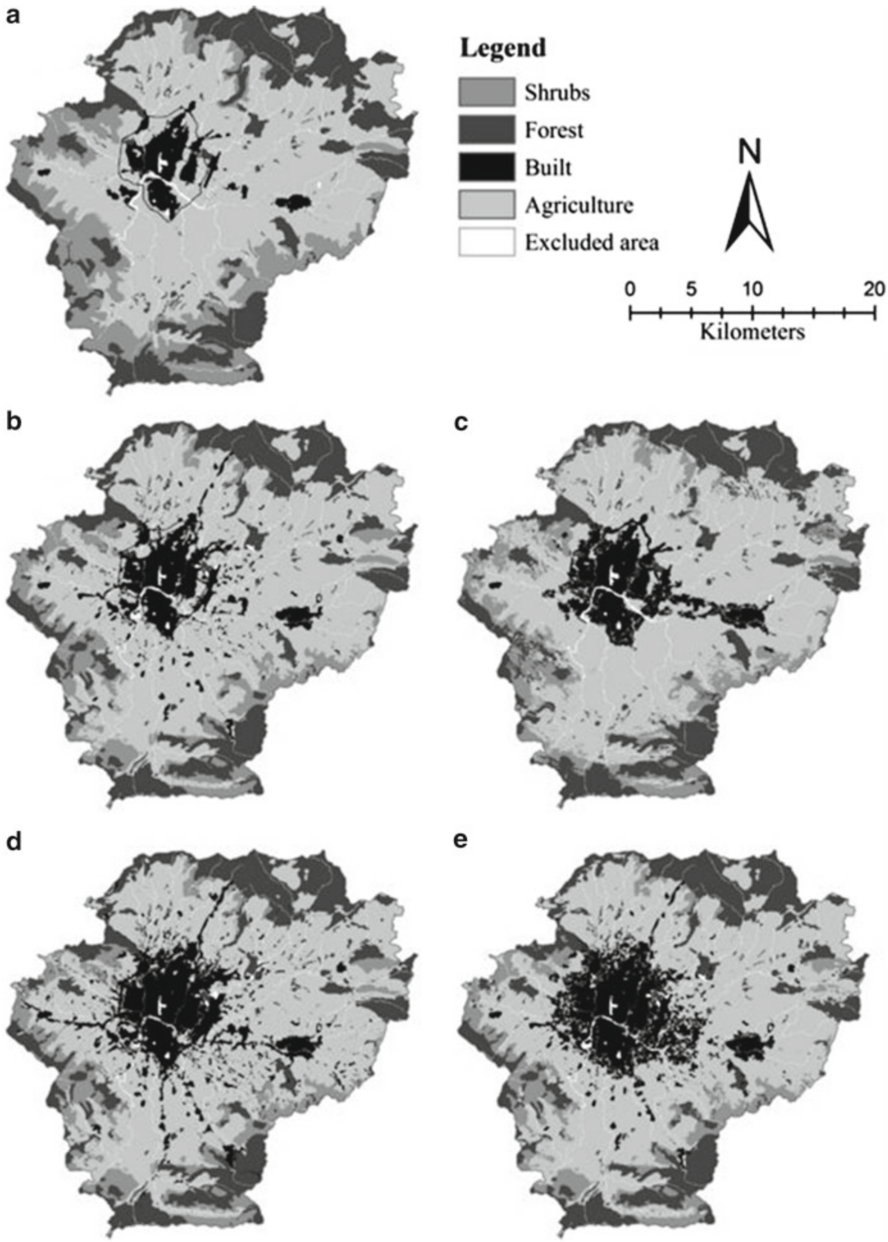


Fig. 13.3 Actual vs. simulated land-use patterns. (**a**) Actual 1978; (**b**) actual 1991; (**c**) simulated 1991; (**d**) actual 2000; (**e**) simulated 2000

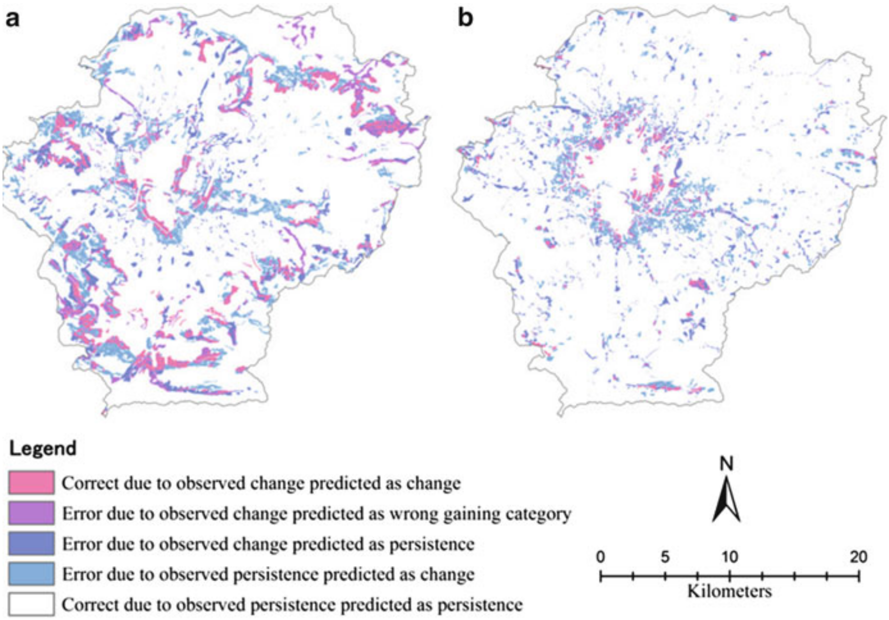


Fig. 13.4 Validation results ((**a**) 1991 and (**b**) 2000) obtained by overlaying the reference map of time 1, the reference map of time 2, and the prediction map of time 2

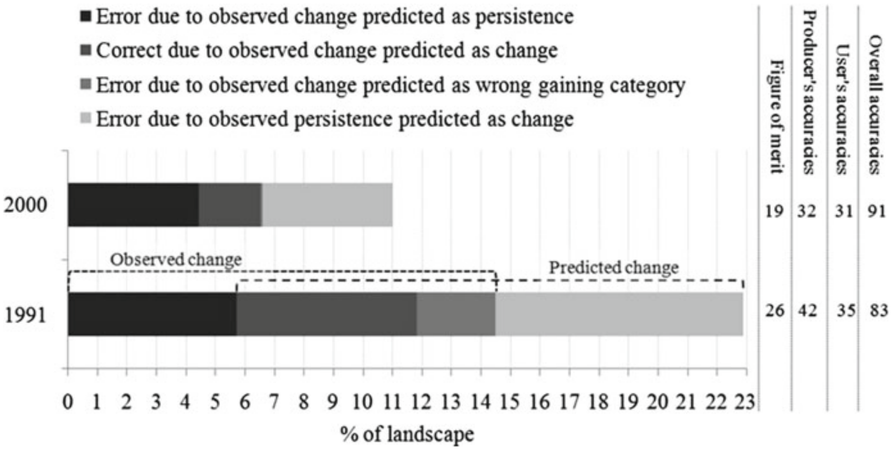


Fig. 13.5 Sources of percentage correct and percentage error in the model validation

Table 13.5 Actual vs. simulated land cover in percentages (1991–2000)

	1991		2000	
Land cover	Actual	Simulated	Actual	Simulated
Shrubs	12.14	12.01	10.71	10.72
Forest	20.78	20.82	20.04	20.10
Built-up	8.58	8.70	13.02	13.02
Agriculture	58.50	58.47	56.23	56.16

center portions of each bar represents the area of change according to the reference maps, and the union of the segments on the center and right portions of each bar represent the area of change according to the prediction map. As represented by the figure of merit, 26% of overlap between the observed change and the predicted change was found in 1991, while this overlap had decreased by 7% in 2000. A decreasing pattern is also noticed in the PA and the UA. However, the overall agreement between the reference and predicted maps, as shown by the OA, was 91% in 2000, which is an increase of 8% compared with 1991. Table 13.5 shows the results by land-cover type at the quantitative level, i.e., actual vs. simulated, with minor differences.

13.3.3 Forecasting Urban Dynamic Patterns for the Years 2010 and 2020

Using the same configuration as the 1991–2000 simulation model, as well as the input map of 2000 (time 1) and the road network of 2000, we performed a simulation aimed at predicting the spatial patterns of urban growth in the metropolitan region for the years 2010 and 2020. Figure 13.6 shows the urban growth consistently expanding eastwards, and encompassing the suburban villages and two urban centers, namely Madhayapur Thimi and Bhaktapur. The built-up surfaces of the villages in the southeastern part were also starting to be overrun by 2010. The current agricultural area between the Madhayapur Thimi and Kathmandu–Lalitpur urban centers will be converted into a built-up area in the 2010s. By 2020, all the urban centers will be aggregated into a greater metropolitan region in the valley.

Figure 13.7 shows the quantitative results produced from the simulated maps by the selected land-cover categories. The built-up area will be increased from 87 km^2 to 148.7 km^2 by 2020, which means a 59% increase, while other land-cover classes are diminishing at different levels. The most vulnerable areas seem to be shrub land changing to either agricultural areas or forest. Most of the agricultural areas will be converted to built-up areas. However, the urban growth rate will decrease in the next few decades (Table 13.6). The trend of negative growth of both forest and agricultural areas is increasing, although at various rates. At the same time, the negative growth of shrub land will be at a lower rate by 2010.

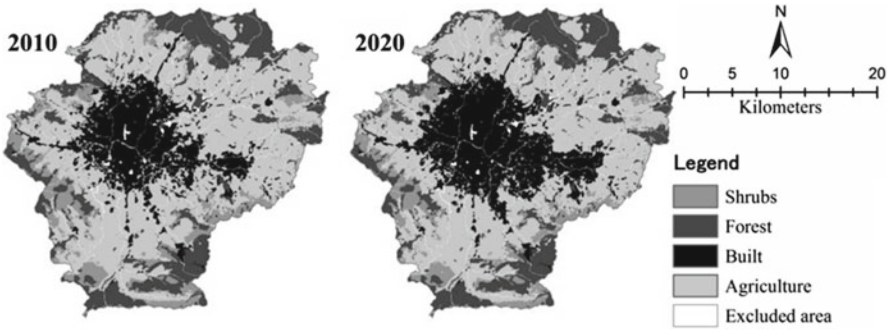


Fig. 13.6 Simulated land-cover patterns (2010–2020)

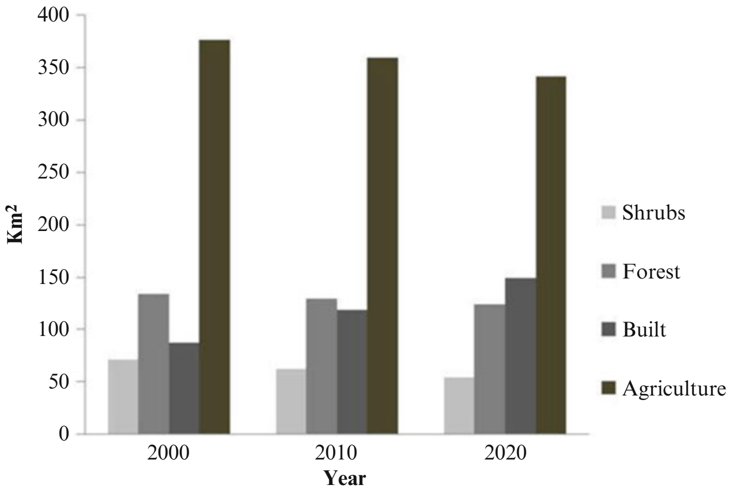


Fig. 13.7 Area of land cover in km^2 (2000–2020)

Table 13.6 Annual urban growth rate[a] (2000–2020)

	Years		
Land cover	2000	2010	2020
Shrubs	−1.298	−1.318	−1.309
		−0.374	−0.381
Built-up	5.755	3.020	2.213
Agriculture	−0.427	−0.473	−0.490

[a] Rate = (((present area−past area)/(past area))×100)/number of past years

13.4 Discussion and Conclusion

Modeling urban growth has been the objective of urban research for many years. This chapter has analyzed the historical land-cover transition and simulated future urban dynamics in the Kathmandu metropolitan region using the Bayesian approach incorporated with CA and GIS approaches. The historical evidence of land-cover transition showed that the rate of encroachment of urban areas on other land-cover areas has been quite rapid, with scattered patches of urban development characterizing the urban sprawl in the metropolitan region. On the one hand, it may have been the conversion of agricultural land to built-up areas in the urban fringes that forced the farmers in the vicinity to migrate. On the other hand, farmers were encouraged to develop agricultural activities in the rural hills owing to road expansion and market accessibility to rural areas, which enhanced deforestation and the encroaching shrub land. After the establishment of democracy in 1990, Kathmandu became the center of political power and the hub of business activities (Thapa et al. 2008). The business and economic opportunities led to a population influx in the valley, creating a demand for housing that eventually increased the built-up surfaces. Landscape fragmentation and heterogeneous land-use development are recent phenomena in the area (Thapa and Murayama 2009; Bhattarai and Conway 2010). This type of development around the city is often described as economically inefficient and aesthetically unattractive (Cadwallader 1996). The form of urban sprawl in uneven spatial directions might have been further promoted by the weakness of the local government regulations, as reported by Thapa et al. (2008). The land tenure system might also have contributed to this form of sprawl, since private land has been subdivided and used for the development of unplanned residential areas.

The modeling framework adopted in DINAMICA, which is used in this research, can be compared to the state change model concept which was first used to model urban development patterns in a CUF II model (Landis 2001). The state change models project future land use and land cover using the information on land use at two moments in time to calibrate a statistical model which relates a set of independent variables to the observed land-use changes at each location (Klosterman and Pettit 2005). Current and projected values for the independent variables are then used to project future land-use changes. A similar concept is applied in DINAMICA differentiating with probability functions, i.e., DINAMICA uses the weight of evidence, while CUF II uses nonordinal logit (Landis 2001). Recently, other models, e.g., a new version of Clarke's urban growth model, SLEUTH-3r (Jantz et al. 2010), and the land change modeler of IDRISI (Eastman 2009) adapted similar state-change concepts by differentiating with the probability function. In order to derive future spatial patterns of urban development in the valley, the calibration of the model was conducted systematically over time, with the similarity of the model's outcomes being compared with the actual urban development of the metropolitan region between 1978 and 2000. The model was applied to generate perspective views of the city for the next 20 years. The results presented dynamic patterns and a composition of Kathmandu's urban development up to the year 2020.

The perspective views of Kathmandu generated by the model highlighted the requirement of urban planning controls, leading to the conservation of more forest and agricultural land and the promotion of intensive development.

This analysis suggests that urban development in the Kathmandu valley will continue through both filling-in existing urban areas and outward expansion toward the east, south, and west directions in the future. Development will be greatly affected by the existing urban space and the transportation network. Unsurprisingly, under extremely unfavorable topographic conditions, close proximity to the road network plays a crucial role in urban development, much more than in cities with gentler topographies and better infrastructure networks (Liu 2009). However, this study assumed that transportation systems would not change during the future simulation period. There are substantial uncertainties in simulating future changes in road systems since they are subject to relatively frequent changes and are often affected by urban transportation policies and land-use planning. A map of the future transportation network in the valley is not available. The topographical constraints on this development will also be important, especially when the urban areas of Kathmandu extend further outward from existing urban areas. The topographical complexity and the unavailability of a future transportation network map could be the main causes of decreasing urban growth rate in the future.

Furthermore, a concrete future land-use plan for the Kathmandu valley is missing in this modeling. Spatial data related to future urban development planning, such as land-use plans, have still not been produced by the authorities. Therefore, the model used in this research to simulate the future growth scenario is missing any planning guidance.

The simulation estimate is based on extrapolation from historic processes which are not guaranteed to continue in the future, but it mirrors spatial patterns of land cover in the metropolitan region if the historic processes do not alter. In this situation, the model has generated maps to show where and how the urban development of Kathmandu is heading in the next two decades from 2000, which may be a critical reference point for decisions guiding the future urban development and land management in the valley. This study has also demonstrated the usefulness of the data acquired from satellite remote sensing and CA-based urban growth modeling in providing land-use and land-cover maps and change information, which are very valuable for planning and research. The approach adopted in this study can be used for the analysis of urban growth and land-cover changes in developing countries where the amount and quality of geographic information and other ancillary data are very limited.

References

Almeida CM, Gleriani JM, Castejon EF, Soares-Filho BS (2008) Using neural networks and cellular automata for modeling intra-urban land-use dynamics. Int J Geogr Inf Sci 22:943–963

Batty M, Couclelis H, Eichen M (1997) Urban systems as cellular automata. Environ Plann B 24:175–192

Bhattarai K, Conway D (2010) Urban vulnerabilities in the Kathmandu valley, Nepal: visualizations of human/hazard interactions. J Geogr Inf Syst 2:63–84
Bonham-Carter G (1994) Geographic information systems for geoscientists: modeling with GIS. Pergamon, New York
Cadwallader MT (1996) Urban geography: an analytical approach. Prentice-Hall, New Jersey
CBS (2001) Population of Nepal (selected data–central development region). His Majesty's Government of Nepal, Kathmandu
Clarke KC, Hoppen S, Gaydos LJ (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco bay area. Environ Plann B 24:247–261
Eastman JR (2009) IDRISI Taiga, Guide to GIS and remote processing. Clark University, Worcester
Godoy MMG, Soares-Filho BS (2008) Modelling intra-urban dynamics in the Savassi neighbourhood, Belo Horizonte city, Brazil. In: Paegelow M, Olmedo MTC (eds) Modelling environmental dynamics. Springer, Berlin, pp 319–338
Goodacre CM, Bonham-Carter GF, Asterberg FP, Wright DF (1993) A statistical analysis of spatial association of seismicity with drainage patterns and magnetic anomalies in western Quebec. Tectonophysics 217:285–305
Haack B (2009) A history and analysis of mapping urban expansion in the Kathmandu valley, Nepal. Cartogr J 46:233–241
ICIMOD/UNEP (2001) Kathmandu valley GIS database. Kathmandu: ICIMOD
Jantz CA, Goetz SJ, Donato D, Claggett P (2010) Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. Comput Environ Urban Syst 34:1–16
Klosterman RE, Pettit CJ (2005) Guest editorial: an update on planning support systems. Environ Plann B 32:477–484
KVUDC (2002) Long term development concept of Kathmandu valley. Kathmandu Valley Urban Development Committee, Kathmandu
Landis J (2001) CUF, CUF II, and CURBA: a family of spatially explicit urban growth and land-use policy simulation models. In: Brail RK, Klosterman RE (eds) Planning support systems: integrating geographic information systems, models and visualization tools. ESRI, Redlands, pp 157–200
Li X, Yeh AG (2002) Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int J Geogr Inf Sci 16:323–343
Liu Y (2009) Modelling urban development with geographical information system and cellular automata. Taylor and Francis, Boca Raton
Pontius RG, Boersma W, Castella J, Clarke KC, de Nijs T, Dietzel C, Duan Z, Fotsing E, Goldstein N, Kok K, Koomen E, Lippitt CD, McConnell W, Sood AM, Pijanowski B, Pithadia S, Sweeney S, Trung TN, Veldkamp AT, Verburg PH (2008) Comparing the input, output, and validation maps for several models of land change. Ann Reg Sci 42:11–47
Portnov BA, Adhikari M, Schwartz M (2007) Urban growth in Nepal: does location matter? Urban Stud 44:915–937
Soares-Filho BS, Alencar A, Nespad D, Cerqueira GC, Dial M, Del C, Solozarno L, Voll E (2004) Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarem-Cuiaba corridor. Glob Chang Biol 10:745–764
Tang J, Wang L, Yao Z (2007) Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. Int J Remote Sens 28:3255–3271
Thapa RB (2009) Spatial process of urbanization in Kathmandu valley, Nepal. PhD Dissertation. Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki
Thapa RB, Murayama Y (2009) Examining spatiotemporal urbanization patterns in Kathmandu valley, Nepal: remote sensing and spatial metrics approaches. Remote Sens 1:534–556
Thapa RB, Murayama Y (2010) Drivers of urban growth in the Kathmandu valley, Nepal: examining the efficacy of the analytic hierarchy process. Appl Geogr 30:70–83
Thapa RB, Murayama Y, Ale S (2008) Kathmandu. Cities 25:45–57
White R, Engelen G (1997) Cellular automata as the basis of integrated dynamic regional modeling. Environ Plann B 24:235–246

Chapter 14
Land Suitability Assessment Using a Fuzzy Multi-Criteria Evaluation*

Duong Dang Khoi and Yuji Murayama

14.1 Introduction

Protected areas (PAs) have become a universally adopted way of conserving biodiversity for a wide range of human values. Globally, 11.2% of the total forest area has been designated for the conservation of biological diversity (FAO 2005). A PA is defined as "an area of land and/or sea especially dedicated to the protection of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means" (IUCN 1994). Although PAs are designed for biodiversity conservation goals, they are also important to the livelihoods of local communities, particularly of indigenous people who depend on the resources available in the PAs for their survival (McNeely 1993; WCPA 2010). In Vietnam, PAs are strongly affected by nearby rural communities because the people's livelihoods often depend heavily on land and forest resources from PAs (Phuong and Dembner 1994; ICEM 2003). Driven by population pressure in the PA buffer zones, increasing demands for food, timber, and non-timber products have resulted in agricultural expansion into PAs (ICEM 2003). To control agricultural expansion

*This chapter has been improved from "Duong Dang Khoi and Yuji Murayama (2010), Delineation of suitable cropland areas using a GIS-based multi-criteria evaluation approach in the Tam Dao National Park region, Vietnam. *Sustainability*, 2, 2024–2043."

D.D. Khoi (✉)
Faculty of Land Administration, Hanoi University of Natural Resources and Environment, Vietnam

Formerly in Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: khoi_tn@yahoo.com

Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: mura1@sakura.cc.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_14,

into PAs and ensure sustainable uses of land in the buffer zones, there is a great need to locate agricultural production activities in suitable locations to avoid negative ecological consequences. PA managers are often requested to identify the spatial distribution of suitable areas for cropland in the buffer zone. One way to achieve this identification is to employ a land suitability assessment (LSA) tool.

An LSA is a prerequisite for determining and locating future land uses (van Ranst et al. 1996; Collins et al. 2001). It is the process of determining the fitness of a given parcel of land for a defined use (Stainer 1991). An LSA involves the selection of the biophysical or socio-economic factors, or both, of an area; the combination of the selected factors with the decision-maker's preferences allows one to create a composite suitability index (Sui 1993). Therefore, it can be conceptualized as a multiple criteria decision-making problem (Pereira and Duckstein 1993). Boolean overlay and modeling approaches such as neural networks and evolutionary algorithms are recently developed methods for performing LSAs in a geographical information science (GIS) environment. However, these approaches lack a well-defined mechanism for incorporating the decision-maker's preferences into the GIS procedures (Malczewski 2006). This disadvantage can be solved by integrating GIS and multi-criteria evaluation (MCE) methods, thus producing an effective tool for multiple criteria decision-making issues (Malczewski 2006). The purpose of MCE is to investigate a number of choice possibilities in the light of multiple criteria and multiple objectives (Cover 1991). An integration of GIS and MCE (GIS–MCE) can help land-use planners and managers to improve decision-making processes (Malczewski 1999). GIS enables the computation of assessment factors, while MCE aggregates them into a land suitability index.

This study aims to delineate the areas suitable for cropland through a GIS-based MCE approach using biophysical factors and the 2007 Landsat ETM$^+$ imagery for the Tam Dao National Park (TDNP) region, Vietnam. We believe that biodiversity conservation efforts can be improved if priority areas for crop farming and sustainable land uses in the buffer zone are modified based on a comprehensive land evaluation. We selected the TDNP region as a case study because this region is the last remaining primary forest near Hanoi, the capital of Vietnam. It contains a rich biodiversity, but several species are known to be threatened by habitat destruction due to agricultural expansion (Khang et al. 2007).

14.2 Methods

14.2.1 Study Area

The TDNP region is one of the most important protected areas in Vietnam. This region is situated in the northern part of Vietnam (Fig. 14.1), and is considered to be one of its best and largest examples of rainforest habitat. It is endowed with a diversity of insects, butterflies, birds, medical plants, and rare animals (Ghazoul 1994). A recent biological survey identified 1,436 plant species and 1,141 animal species (Khang et al. 2007).

The region is characterized by a tropical monsoon climate with a mean annual rainfall of around 2,600 mm; most of the rainfall occurs from April to October. The

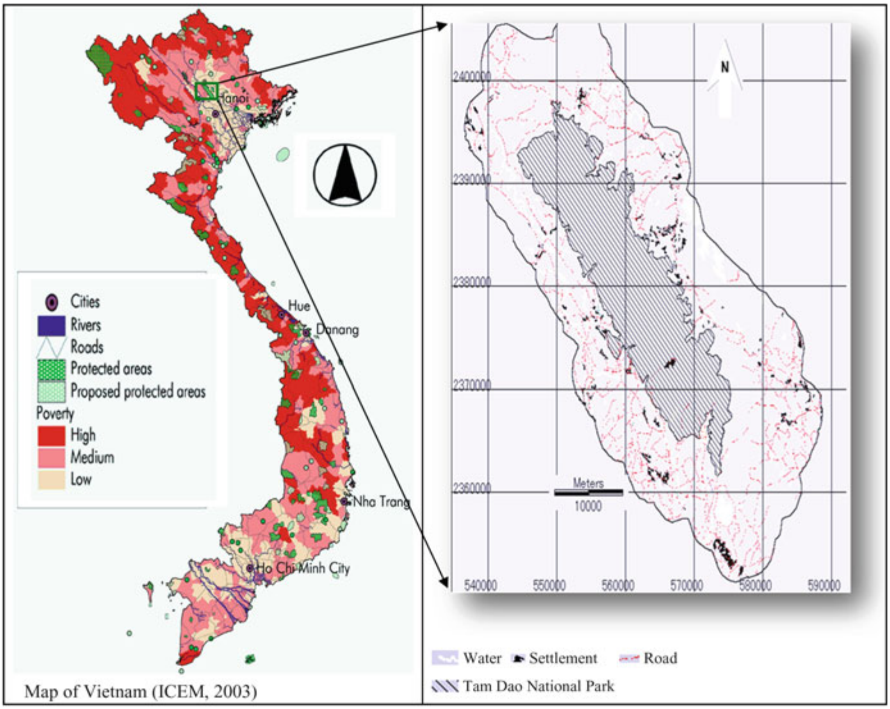


Fig. 14.1 Location of the Tam Dao National Park region, Vietnam

terrain of the area is mostly undulating with steep pediments, and the elevation ranges from 100 to 1,580 m above mean sea level. The total study area is spread over 141,328 ha, which includes the TDNP (35,000 ha) and the buffer zone. The current land uses over the entire study area are primary forest, secondary forest, rain-fed agriculture, paddy rice, settlement, and water. Deforestation from illegal logging and agricultural expansion has been causing serious land degradation (Fig. 14.2) because most of the 200,000 people living in the buffer zone of the TDNP generate their incomes from small-scale farming. Aside from agricultural production, few economic activities exist in the area (TDMP 2005). Currently, the main challenge for the buffer zone is to ensure food production for the growing population while supporting biodiversity conservation goals. The cropland was chosen for this investigation because the expansion of cropland influences the sustainability of the TDNP.

14.2.2 Input Data and Landsat Image Processing

The input data used for this study were based on the selected evaluation factors discussed in the next section. They include a topographical map, soil map, water resource map, road network map, and park boundary map (Table 14.1). These data

Fig. 14.2 Forest logging for agricultural expansion in the buffer zone of the TDNP (photograph by the author, 2009)

Table 14.1 List of databases used in this research

Data types	Year	Scale/resolution	Sources
Topographical map	1972	1:50,000	TDNP Management Office
Soil map	2005	1:100,000	National Institute for Agricultural Planning, Ministry of Agriculture and Rural Development, and TDNP Management Office
Road network	2007	–	TDNP Management Office
Water bodies	2007	–	TDNP Management Office
Park boundary	2007	–	TDNP Management Office
Landsat images	2007	28.5 m	University of Maryland
Field survey	2009.3	–	TDNP region

were used for delineating areas suitable for cropland. Landsat satellite images were used to derive the current land-use map to analyze spatial matching between the current land uses and suitability patterns.

Once the databases were collected, thematic maps were developed for each factor. A digital elevation model (DEM) was constructed using a contour map with a scale of 1:50,000 and an interval of 20 m. The slope factor was derived from this DEM. Soil texture, soil depth, soil organic matter, and soil pH factor maps were extracted from the digitized soil map with a scale of 1:100,000. The distances to water, roads, and the park boundary were generated from the water, road network, and park boundary maps, respectively. The resolution of all raster factor maps was set at 30 m × 30 m.

The Landsat satellite images acquired in 2007 were used to derive the recent land-use map. Six bands (bands 1–5 and 7) were processed to derive the land-use map. The image was rectified to a common UTM/WGS84 coordinate system that is based

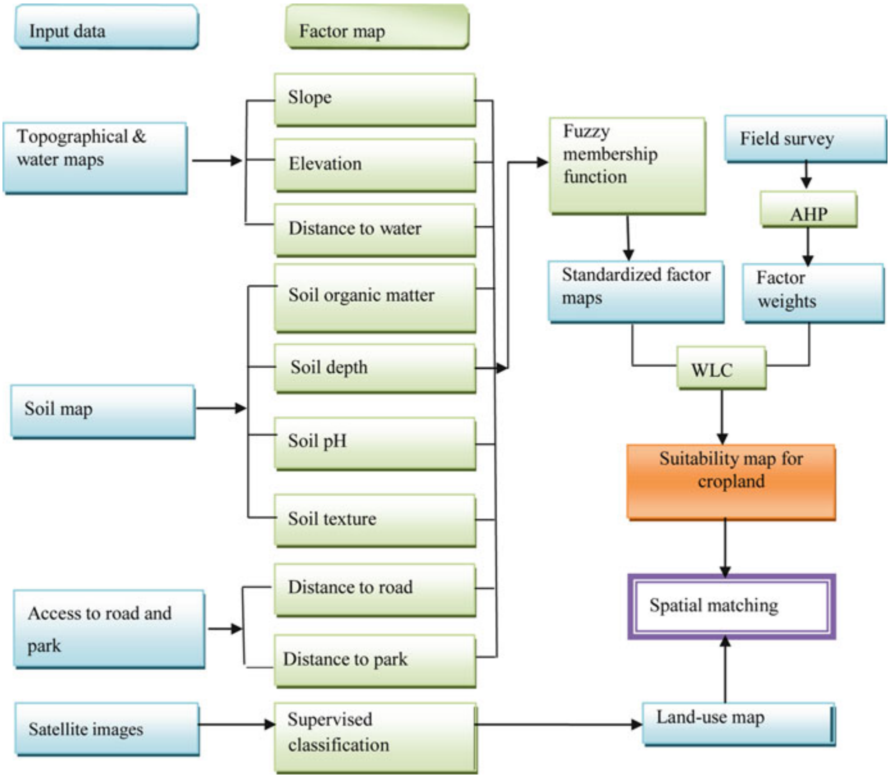


Fig. 14.3 Flowchart of the land suitability assessment for cropland (*AHP* analytical hierarchy process, *WLC* weighted linear combination)

on the topographic map. The clusters of pixels representing various land-use types were identified as training sites that are based on unsupervised classification, an existing land-use map, and the knowledge of the authors on the relative locations of land-use types. After all the training sites were identified and digitized by the on-screen method, the class signatures were generated. A maximum likelihood method was used to classify these images into the land-use map. The accuracy of the classified map was investigated. A stratified random sampling design was employed to identify 270 locations (pixels) for field data collection. During the field trip, GPS equipment was used to trace geographical data, and a digital camera was used to record views of the locations for laboratory analysis.

14.2.3 Multi-Criteria Evaluation

The GIS–MCE procedure for the cropland suitability assessment in the TDNP region included several stages that are shown in Fig. 14.3. A determination of the relevant factors was the first step in the assessment, and this was followed by standardizing and weighting the factors, combining them with their weights, and finally spatially matching the suitability map and the current land-use map. The procedures and algorithms available in IDRISI Taiga (Eastman 2009) were employed to implement the assessment.

Initially, the factors were selected based on their relevance to the suitability of cropland and the availability of databases. The selection of factors is a technical process that is based on expert knowledge or empirical research. We selected 12 experts to be involved in the assessment, who were all between 30 and 50 years of age. They participated in selecting the factors, identifying the suitable ranges of the factors, and evaluating the weights of the factors. They included five agronomy experts, five soil experts, and two forestry experts. Eleven of the experts had bachelor's degrees, and one had a master's degree. These experts have all worked for at least 5 years at the office of the TDNP, and have also worked for the district department of agriculture and rural development in the region. After discussion with the experts during the field survey period, nine factors (slope, elevation, distance to water, soil organic matter, soil depth, soil pH, soil texture, distance to roads, and distance to the TDNP boundary) were identified as being the most relevant for the suitability assessment of crop-growing areas in the region. The elevation, slope (terrain), and the distance to water are important determinants of cropland suitability because the terrain often has a relationship with soil fertility as well as with the vulnerability to soil degradation. The slope relates to the retention and movement of soil particles and the rates of runoff and soil erosion; therefore, it closely regulates the soil quality condition. The soil characteristics (soil organic matter, soil depth, soil pH, and soil texture) represent the soil nutrients and water availability for crop growth. The distance to roads is important for crop production because it relates to the transportation costs of input and output items. The distance to the park is defined as the suitability, which is monotonically reduced in areas closer to the park boundary. This variable is included in the LSA because cultivation areas closer to the park may alter the environmental quality of the protected area more seriously.

As the factor maps were originally measured in different scales, they have to be standardized to a uniform suitability rating scale. The MCE method used requires that all factors must be standardized. The standardization transforms the disparate measurement units of the factor maps into comparable suitability values (Eastman 2009). The fuzzy membership function (FMF) approach was applied to standardize the factors. This method provides a useful means of dealing with the uncertainty that results from imprecise boundaries between suitability classes (McBratney and Odeh 1997; Ahamed et al. 2000). An FMF is characterized by a fuzzy membership grade that ranges from 0 (non-membership) to 1 (complete membership) (Eastman 2009). For each factor, the least suitable level was defined as 0, and the most suitable level was defined as 1. Several FMFs can be used to standardize the factors. The sigmoidal FMF is one of the FMFs most widely used in land evaluation (Eastman 2009). In this study, a sigmoidal monotonically decreasing FMF (SMDFM) and a sigmoidal monotonically increasing FMF (SMIFM) were employed. Higher values of elevation, slope, distance to water, and distance to roads would indicate continuously decreasing

Table 14.2 Suitable ranges used for the fuzzy membership function

Factor	Non-membership (unsuitable)	Membership grade (suitable range)	References
Slope (°)	>15	1–15	TDNP agronomy experts Slope from 1° to 25° (Liu et al. 2006), 1° to 15° (Quan et al. 2007; Wang et al. 2007)
Elevation (m)	>400	1–400	TDNP agronomy experts Elevation from 1 to 500 m (Quan et al. 2007)
Distance to water (m)	>2,000	100–2,000	TDNP agronomy experts
Soil organic matter (%)	<0.5	0.5–2.3	TDNP agronomy experts Less than 1–3% (Quan et al. 2007)
Soil depth (cm)	<20	20–150	TDNP agronomy experts Soil depth range from 10 to 60 cm (Quan et al. 2007), 15 cm to more than 30 cm (Wang et al. 2007)
Soil pH	<4.5 and >7.5	4.5–6.9	TDNP agronomy experts pH range from 5 to 8 (Quan et al. 2007)
Soil texture (class)	–	Sandy clay loam, sandy loam, silt loam, loam	TDNP agronomy experts Medium loam is most suitable, light and heavy loam is moderately suitable, sandy loam and medium clay is marginally suitable (Quan et al. 2007; Wang et al. 2007)
Distance to roads (m)	>4,000	100–4,000	TDNP agronomy experts
Distance to the park boundary (m)	<500	500–11 277	TDNP agronomy experts

suitability, and therefore the SMDFM was used to standardize these factors. On the other hand, higher values in the factors of soil organic matter, soil depth, soil pH, and distance to the park boundary would show continuously increasing suitability, and thus the SMIFM was used to standardize these factors. Suitable values for soil texture were assigned according to each textural class. Detailed descriptions of sigmoidal FMFs can be found in Eastman (2009). To apply the FMF, suitable ranges of the factors that define the least and greatest suitability levels were determined based on the experts' knowledge, and somewhat similar studies have been successfully conducted for cropland suitability assessment (Liu et al. 2006; Quan et al. 2007; Wang et al. 2007) (Table 14.2). Suitable ranges for the factors were identified according to the opinions of the experts, and they were also verified by our field visits. Figure 14.4 shows the results of the standardized factor maps. A standardized factor map consists of pixels with continuous scores varying from 0 to 1. A higher pixel score indicates a higher suitability level for that pixel.

The evaluation of suitability involves many factors, and each should be weighed according to its relative importance for the growth conditions of crops. The weight

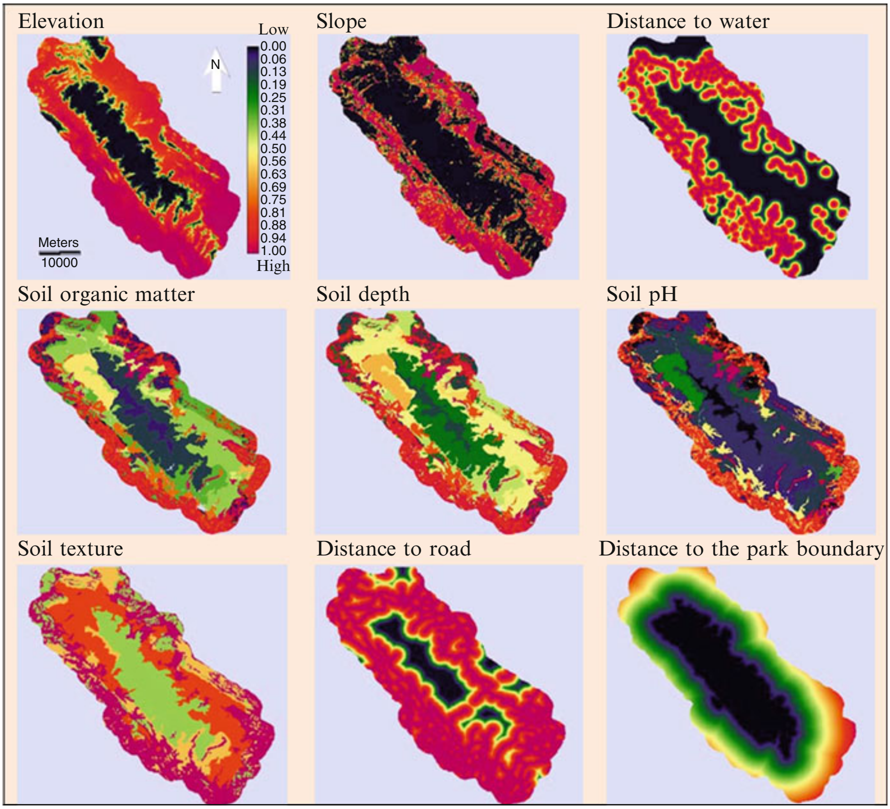


Fig. 14.4 Standardized factor maps (the legend is the same as that for the elevation map for all factor maps)

of each factor was estimated from a pair-wise comparison matrix (PWCM) constructed according to a pair-wise comparison method (PCM) (Table 14.3). The PCM developed by references (Saaty 1980; 1990), in the context of a decision-making process known as the analytic hierarchy process, is the most commonly used method (Eastman et al. 1995). In the PWCM, a pair-wise comparison is a rating of the relative importance of two factors regarding the suitability of the cropland. The PWCM method uses a scale with values from 9 to 1/9 to rate the relative importance of the two factors. A rating of 9 indicates that in relation to the column factor, the row factor is more important. On the other hand, a rating of 1/9 indicates that in relation to the column factor, the row factor is less important. In cases where the column and row factors are equally important, they have a rating value of 1.

In determining the ratings, the 12 experts previously mentioned worked as a group to determine the ratings of the factors. To reach agreement in rating the relative importance of the factors, a majority rule was applied. This means that each rating in the PWCM was compared and decided on based on the agreement of the

Table 14.3 The pair-wise comparison matrix for evaluating the relative importance of the factors for each land-use requirement (the number indicates the rating of the row factor relative to the column factor)

Terrain and water	Slope	Elevation	Distance to water	Weight
Slope	1	1	2/3	0.2856
Elevation	1	1	2/3	0.2856
Distance to water	3/2	3/2	1	0.4288
Consistency ratio (CR) = 0.000				

Soil quality	Soil organic matter	Soil depth	Soil pH	Soil texture	Weight
Soil organic matter	1	3	2	3/2	0.4073
Soil depth	1/3	1	3/2	2	0.2384
Soil pH	1/2	2/3	1	1/2	0.1444
Soil texture	2/3	1/2	2	1	0.2099
CR = 0.087					

Access to roads and the park	Distance to road	Distance to the park	Weight
Distance to roads	1	3/2	0.6000
Distance to the park	2/3	1	0.4000
CR = 0.000			

Land-use requirement for the assessment of site suitability for cropland	Terrain and water	Soil quality	Access to roads and the park	Weight
Terrain and water	1	1/2	3	0.3338
Soil quality	2	1	3	0.5247
Access to roads and the park	1/3	1/3	1	0.1415
CR = 0.046				

majority of experts. In the context of the workshop for determining the relative importance of the factors, a description of the evaluation purpose, an identification of the set of relevant factors, and an explanation of a PWCM and completion procedure were carried out. After careful examination and discussion of the set of factors, the group made all the pair-wise comparisons for the set of factors. The PWCMs developed are shown in Table 14.3. The weights of the factors were then calculated from these PWCMs. The consistency ratios (CRs) of 0.000–0.087 in the table were within acceptable levels (Saaty 1980, 1990). According to Saaty (1980, 1990), the calculated CR must be less than 0.1, which is the acceptability cut-off point. This means that if the computed CR is less than 0.1, the calculated weights of the factors are consistent. If the calculated CR is more than 0.1, the PWCM needs to be re-evaluated, and the weights of the factors also need to be re-calculated accordingly. An example of spreadsheet calculations for the CR of overall site suitability factors for cropland is shown in Table 14.4. The points (a) and (b) show the calculation of the factor weights. Parts (c)–(e) show the calculations of the CR.

Table 14.4 Example of spreadsheet calculations for the consistency ratio of site suitability for cropland

	Values			Decimal			Normalization							
	TW	SQ	RP	TW	SQ	RP	TW	SQ	RP	Weight	λ	CI	RI	CR
TW	1	1/2	3	1.00	0.50	3.00	0.30	0.27	0.43	0.3338				
SQ	2	1	3	2.00	1.00	3.00	0.60	0.55	0.43	0.5247	3.0538	0.0269	0.58	0.0464
RP	1/3	1/3	1	0.33	0.33	1.00	0.10	0.18	0.14	0.1415				
Sum				3.33	1.83	7.00				1.0000				

TW terrain and water, *SQ* soil quality, *RP* access to roads and the park

(a) Sum the numbers in each column of the values matrix; divide each number in the decimal matrix by the column sum; the resulting matrix is the normalization matrix.
(b) Take the average of the numbers in each row of the normalization matrix; the average value is the weight.
(c) Compute *lambda* (λ) by the following steps (Malczewski 1999):

1. Determine the weighted sum vector by multiplying the weight of the TW, the weight of the SQ, and the weight of the WP by the first column, the second column, and the third column of the values matrix, respectively, and finally, sum these values over the rows;
2. Determine the consistency vector by dividing the weighted sum vector by the factor weights as follows:

Step 1	Step 2
(1) (0.3338) + (0.5) (0.5247) + (3) (0.1415) = 1.0208	1.0208/0.3338 = 3.05837
(2) (0.3338) + (1) (0.5247) + (3) (0.1415) = 1.6169	1.6169/0.5247 = 3.08168
(0.3333) (0.3338) + (0.3333) (0.5247) + (1) (0.1415) = 0.4277	0.4277/0.1415 = 3.02140

Then, $\lambda = (3.05837 + 3.08168 + 3.02140)/3 = 3.0538$

(d) The consistency index (CI) is $(\lambda - n)/(n-1)$, $(3.0538-3)/2 = 0.0269$.
(e) The consistency ratio (CR) is CI/RI, where RI is the random consistency index. For $n = 3$, RI = 0.58 (Saaty 1980).
CR = 0.0269/0.58 = 0.0464.

After the standardized factor maps and the weights of the factors had been constructed and generated, the weighted linear combination (WLC) was used to combine the standardized factors and their corresponding weights to obtain an overall suitability map for the cropland (Eastman et al. 1995). All of the factors were combined as $\mathrm{Grid}_{\mathrm{result}} = \sum(\mathrm{Grid}_i \times \mathrm{Weight}_i)$, where Grid_i is the factor i, and Weight_i is the relative weight of factor i. Specifically, the three factors of terrain and water, the four factors of soil quality, and the two factors of access to roads and the park were calculated by (14.1)–(14.3), and then they were all overlaid to produce the overall cropland suitability map according to (14.4). Finally, the recent land-use map and the suitability map were overlaid to analyze the spatial matching. A simple overlay technique was used between the land-use map and the suitability map, and then the statistics of the suitability classes for each land use were calculated.

$$\text{Terrain and water grid} = \text{Grid}_{\text{slope}} \times 0.2856 + \text{Grid}_{\text{elevation}} \times 0.2856 + \text{Grid}_{\text{distance to water}} \times 0.4288 \tag{14.1}$$

$$\text{Soil quality grid} = \text{Grid}_{\text{soil organic matter}} \times 0.4073 + \text{Grid}_{\text{soil depth}} \times 0.2384 + \text{Grid}_{\text{soil pH}} \times 0.1444 + \text{Grid}_{\text{soil texture}} \times 0.2099 \tag{14.2}$$

$$\text{Access to roads and the park grid} = \text{Grid}_{\text{distance to roads}} \times 0.6 + \text{Grid}_{\text{distance to the park}} \times 0.4 \tag{14.3}$$

$$\text{Overall suitability grid} = \text{Grid}_{\text{terrain and water}} \times 0.3338 + \text{Grid}_{\text{soil quality}} \times 0.5247 + \text{Grid}_{\text{access to roads and the park}} \times 0.1415 \tag{14.4}$$

14.3 Results

Figure 14.5 shows the suitability map for the cropland in the TDNP region. The map contains pixels with varying degrees of suitability from 0 to 1. A higher pixel score shows a higher suitability level. For easier representation, the map was re-classified into four classes based on the structure of the FAO suitability classification (FAO 1976): most suitable (0.75–0.96), moderately suitable (0.5–0.75), marginally suitable (0.25–0.5), and least suitable (0–0.25). The most suitable is the land with minor limitations that do not significantly affect crop farming. The moderately suitable is the land with limitations that, in aggregate, are moderately limiting to crop farming. The marginally suitable is the land that has limitations which, in aggregate, are severely damaging to crop farming. The least suitable is the land with limitations that, in aggregate, are very severely damaging to crop farming. The extent of each class is summarized in Table 14.5. The result indicates that 28.10% of the total study area was found to belong to the most suitable class. These most suitable areas are mainly characterized by flatness, a nearness to water, and deep soil depth. The moderately suitable class was found to make up 23.96% of the territory. The most and moderately suitable classes together comprise 52.06% of the total area, whereas the existing cropland area was 46.5%. This result highlights that the most and the moderately suitable areas have been used for cropland in the region. The least suitable and marginally suitable classes make up 19.17% and 28.77%, respectively. These areas are often located in areas with steep terrain, low soil depth, and less water access. If farmers are forced to reclaim land for agriculture due to population pressures, the marginally suitable areas that are highly vulnerable to soil erosion may be the target areas of the future.

Different factors have different importance levels for the site suitability of cropland. The result of evaluating the relative importance of different factors shows that the soil quality (soil organic matter, soil depth, soil pH, and soil texture) is the most

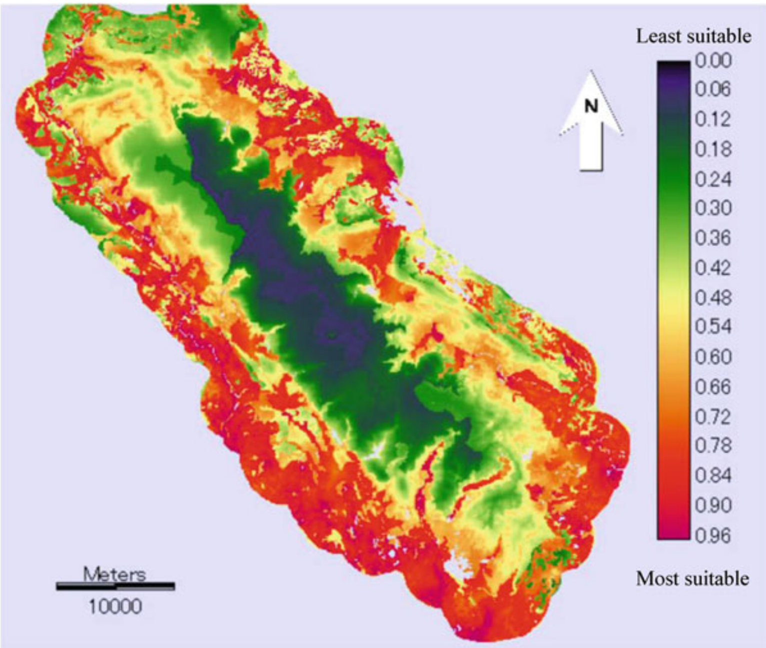


Fig. 14.5 Land suitability map for cropland in the TDNP region

Table 14.5 Area of cropland suitability classes

Suitability class	Area (ha)	Proportion (%)
Least suitable	27,069	19.17
Marginally suitable	40,639	28.77
Moderately suitable	33,846	23.96
Most suitable	39,683	28.10

important, followed by the terrain and water (slope, elevation, and distance to water), and access to roads and the park (distance to roads and distance to the park boundary). Soil quality, with a weight of 0.5247, is determined to have a major impact on the overall suitability because it regulates the storage of soil nutrients and the water-holding capacity, which are necessary biophysical conditions for crop growth. The topographical and water factor, with a weight of 0.3338, is the second contributor. The slope affects the retention and movement of water and soil particles, the runoff rate, and accelerated soil erosion. These effects are closely linked to the soil quality conditions. Elevation relates to increased water-pumping costs for agricultural production. Water availability is very important for crop growing in the area. Natural lakes, ponds, streams, and rivers are major water providers for agricultural production in the area. Water resources in the region mostly depend on sources from the

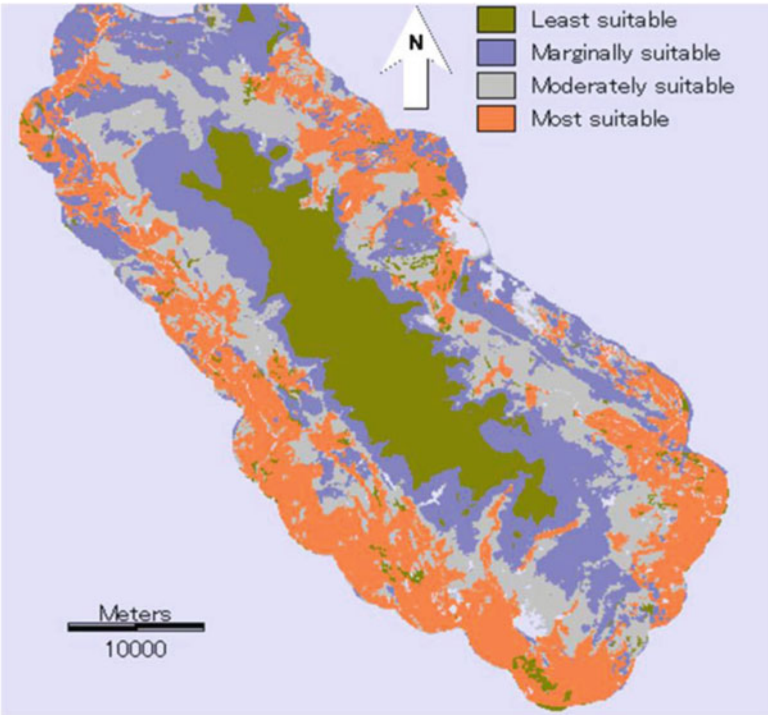


Fig. 14.6 Map of the suitability zones

TDNP forest ecosystems. Therefore, there is a strong link between the conservation of native forest ecosystems and agricultural development in the region. The access to roads and the park plays a weaker role compared to the others. Road networks are significant for local communities because they enhance commercial agricultural activities and transportation. The distance to the park boundary affects the biodiversity conservation activity of the TDNP; therefore, it relates to the site suitability of cropland.

The spatial matching offered valuable information for identifying whether the land was optimally utilized in the region. The result of overlaying the suitability map (Fig. 14.6) with the land-use map of 2007 (Fig. 14.7) is given in Table 14.6. The accuracy of the land-use map based on Kappa statistics was 90.1%. The land-use map indicates that the major land uses are primary forest (25,459 ha), secondary forest (44,018 ha), rain-fed agriculture (41,117 ha), paddy rice (24,567 ha), settlement (3,130 ha), and water (2,947 ha), which account for 18.03%, 31.17%, 29.11%, 17.39%, 2.22%, and 2.09%, respectively, of the total study area. The primary forest, which is mainly dense native vegetation, is mainly located in the park. The secondary forest includes both forest plantations and shrubs. The rain-fed agriculture is characterized by a mixture of crops, mainly soybeans, peanuts, vegetables, and

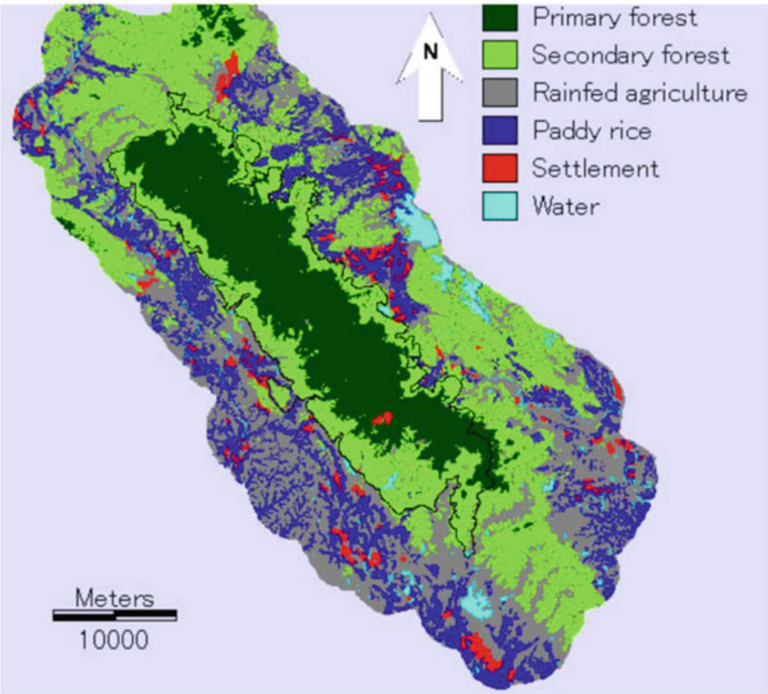


Fig. 14.7 The land-use map derived from the Landsat images of 2007

maize. Paddy fields are used only for rice production. Settlements consist of small houses, front- and backyards, and home gardens. Water resources include a variety of natural lakes, ponds, streams, and rivers.

As expected, the most suitable and moderately suitable areas were found in the existing rain-fed agriculture and paddy fields. The result indicates that 95.22% of the most suitable class was distributed over the rain-fed agriculture and the paddy rice, while only 3.37% of the class was located in the secondary forest. With respect to the moderately suitable class, 83.01% of the class was found in the rain-fed agriculture and the secondary forest, whereas only 15.94% of the class was located in the paddy rice. For the marginally suitable class, 71.37% of the class was found in the secondary forest. This class was also found in the primary forest (14.94%) and the rain-fed agriculture (12.55%). Finally, the least suitable class was mainly stretched over the primary forest. The most and moderately suitable areas have already been utilized for paddy rice and rain-fed agricultural crops. Although some of the rain-fed agricultural areas may cause land degradation due to soil erosion, these utilized lands may not be easy to change to more sustainable uses, such as agro-forest farming or fruit trees, in the future because of the growing population in the area.

Table 14.6 The overlaying result between the suitability map and the land-use map of 2007

	Level of suitability								
	Most suitable		Moderately suitable		Marginally suitable		Least suitable		Total land-use
Land-use type	ha	%	ha	%	ha	%	ha	%	ha
Primary forest	8	0.02	42	0.12	6,072	14.94	19,337	71.44	25,459
Secondary forest	1,339	3.37	11,359	33.56	29,006	71.37	2,314	8.55	44,018
Rain-fed agriculture	19,039	47.98	16,738	49.45	5,099	12.55	241	0.89	41,117
Paddy rice	18,748	47.24	5,394	15.94	251	0.62	174	0.64	24,567
Settlement	549	1.38	313	0.92	212	0.52	2,056	7.60	3,130
Water	–	–	–	–	–	–	2,947	10.89	2,947
Total suitable class	39,683	100.00	33,846	100.00	40,640	100.00	27,069	100.00	141,238

It is important to note that the farmers are not aware of formal LSA methods, but instead trust their own experience regarding land suitability. The farmers have a profound knowledge of their lands, and classify the suitability of the land according to crop yield. Crop yield often correlates with the biophysical factors of the soil, such as terrain, fertility, and water availability. In the study, the factors selected based on the opinions of the experts cover the farmers' perceptions; therefore, the assessment spatially matches the majority of existing cropland.

14.4 Discussion and Conclusion

In general, developing countries have adopted the North American approach to the management of PAs. This approach only emphasizes nature conservation (Colchester 1997), but the livelihoods of local populations in nearby PAs have often been ignored (Pimbert and Pretty 1997). The recognition of PAs as part of a broader socio-economic system (McNeely 1993) led to the concept of protected area-buffer zone land-use management. The maintenance of the local communities' livelihoods and conservation has been challenging the TDNP managers. It argues that sustainable land use may not be achieved in isolation because agricultural expansion and poverty are interrelated. Poverty is a primary cause of cultivation of large areas of sloping lands in the region, which is exacerbating land degradation. Therefore, land-use policy and other development policies should be combined to improve the living standards of the poor, and thereby possibly reduce the over-exploitation of land resources and land degradation in the area.

An LSA is a preliminary stage for assessing whether land is likely to be practical and successful for sustainable development of the intended goals. In many cases, cropland has been promoted in areas that are unsuitable in terms of soil conditions. Owing to increasing population pressures, agricultural expansion has been increasing in the TDNP region without consideration of the suitability of the site. To ensure sustainable land uses, there is a great need to allocate farming activities to suitable locations in order to prevent undesirable effects on biodiversity conservation efforts as well as land degradation in the region. The TDNP and its buffer zone management documents were reviewed. The TDNP-buffer zone management emphasizes the integrity of the buffer zone and the TDNP, particularly regarding sustainable land uses in the buffer zone (Khang et al. 2007; TDMP 2005). Buffer zone land-use management affects the protected area because the environmental quality of the buffer zone is critical to maintaining the ecological functions of the protected area (Bridle et al. 2004). In the TDNP region, land is an important resource for the enhancement of the living standards of local people near the PA. Land-use management of the buffer zone is facing the issue of balancing agricultural development and forest conservation. The approach and results presented in this study may support land-use management decisions towards a more sustainable PA system. It is assumed that different crop farming strategies should be practiced according to the varying degree of suitability. Therefore, we recommend farming strategies for four zones according to the four levels of suitability shown in the land suitability map. Sustainable land-use projects supported by local and central governments should receive investment in accordance with each of the zones. We believe that if such farming strategies are introduced according to these zones, they can prevent further deforestation and improve the appropriate use of land in the buffer zone.

First, for the most suitable zone, most of the zone has been used for rain-fed agriculture and paddy rice. This zone is distributed over the lowlands around the region (Fig. 14.6). Every household can improve their income if the productivity of crops in this zone is improved. Therefore, a greater intensification of crops such as paddy rice and maize should be encouraged in order to enhance agricultural productivity in this zone, and thus the production pressure on the marginally suitable zone can be reduced. This strategy has been successfully implemented, and it can arguably be explained as one of main causes for the increase in reforestation across Vietnam (Meyfroidt and Lambin 2008). However, some of the secondary forest (1,339 ha) was evaluated as highly suitable for cropland. These areas may be converted into cropland owing to population pressures. In such situations, agro-forestry systems (AFS) should be practiced. AFS, the combined use of crops and trees on the same area of land, plays ecological, social, and economic roles (Jianbo 2006). For example, AFS can reduce soil erosion and the loss of soil nutrients, improve landscape diversity (Palma et al. 2007), and generate income for farmers. If an AFS is practiced, a state subsidy for farmers may be needed in the long-term because this system often generates less income (Khang et al. 2007). The subsidy can be understood as a means of compensation for farmers because they contribute to conservation efforts through more sustainable land use. The state subsidy should become a common policy for all protected areas across the country.

Second, for the moderately suitable zone, the majority of the zone has also been used for rain-fed agriculture and paddy rice. The diversification of crops and AFS may be a strategic option for the moderately suitable zone. The current least-productive rain-fed agriculture should be converted into perennial crops such as tea and fruit trees. These crops can increase land coverage and thus can be a more sustainable land-use type. However, a large portion of the secondary forest was assessed to be moderately suitable for cropland (Table 14.6). This portion can potentially be converted into cropland. The conversion of this portion into cropland should be restricted because it is spread over steep land that is highly vulnerable to soil erosion. Third, the marginally suitable and unsuitable zones should not be used for agricultural activities because most of these zones are the primary and secondary forest. This restriction may be linked to the reduced welfare of the population. This requires state support for people who are heavily dependent on natural resources for their well-being. For example, non-farming jobs, such as handicrafts and ecotourism, can be alternative or additional livelihoods that should be considered.

Some concrete measures should be considered for the implementation of agricultural intensification and AFSs which prevent further deforestation and land degradation in the region. It is argued that intensification and AFSs are the key activities that enhance sustainable land use. Crop intensification systems may minimize the expansion of new cultivation areas into the forest because they are expected to increase agricultural productivity in the region. Measures for the intensification systems should aim at enhancing the local farmer's capacity via support projects. These projects should focus on the irrigation system, hybrid crop varieties, soil nutrient management, and integrated pest management. Improvements in the irrigation system can trigger an increase in maize area in the winter season, and thus agricultural productivity would be substantially increased. The introduction of hybrid crop varieties in combination with soil nutrient management and integrated pest management can be implemented through technical training courses for the farmers. For the sloping lands, there is a definite need to combine forestry, crop, and animal husbandry on individual farms to replace mono-cropping. These combinations can have a synergistic effect on the productivity of the land and its resilience to degradation. The AFS should be demonstrated in the first introductory step because these systems are not commonly practiced by the farmers in the region. The purpose of the demonstration would be to help the local farmers to acquire knowledge of how to use their sloping land efficiently. Both technical and financial support is very important for the implementation of the systems. Agricultural extension workers should be employed for each commune or village, because timely technical support for the farmers is needed.

It should be noted that the land suitability map is intended to guide regional land-use decisions. From the TDNP management perspective, the map can be used for a decision-making process that allocates land to the uses that provide the greatest benefits to the conservation of biodiversity and other ecosystem services. However, on a local scale, the farmers may not adopt the land-use plan because their land-use decisions are affected by several other factors: mainly their economic conditions. Most farmers avoid bank loans if they have other investment choices. They commonly adopt low-investment alternatives because self-subsistence farmers tend to adopt short-term objectives in nature, and they are likely to give a low priority to long-term benefits. This low investment involves an enlargement of the cultivation area, which goes against conservation or the maintenance of ecosystem services. The managers are confronted with the need to make difficult decisions. This may require a negotiation process between the managers and the farmers. Once a consensus has been gained through negotiation, the practicality of the plan's implementation can be achieved.

Although the GIS–MCE approach provides an effective framework for land evaluation, the selection of assessment factors and the identification of a suitable range for each factor have a direct influence on the results. In this study, the factors were selected based on the knowledge of local experts; therefore, they represent a considerable share of the factors relevant to the suitability of growing areas in the region. In addition, the FMF approach was used to standardize the factors. The FMF approach is useful because it is good at dealing with land-use suitability classes that do not have clearly defined boundaries (Groenemans et al. 1997). Therefore, the suitability map represents a more accurate result. In particular, the integration of spatial databases and expert knowledge significantly enhances decision-making capacity when undertaking land suitability evaluations. Moreover, the approach highlights participatory decision-making processes (Eastman et al. 1992). Therefore, it can minimize and solve conflicts among competing interests in the area of protected area–buffer zone land-use management.

The GIS–MCE approach has been widely applied in land suitability analysis (Malczewski 2006). However, the application of this method in protected area–buffer zone management is relatively new in Vietnam. As a tool for decision support, GIS–MCE has shown a capacity for making choices among land-use alternatives. The MCE of soil, topography, and accessibility factors was exemplified to be useful for delineating areas suitable for cropland in the TDNP region. In particular, the involvement of local experts was vital to obtaining consistent results. The experts played key roles in the selection of the evaluation factors and in the determination of the factor weights. The remote-sensing data offered land-use information that was crucial to examining the spatial matching between the potential suitability areas and the current land-use patterns. This information helped to identify whether the land has been used optimally, and whether future land uses can be modified for the region. The application presented in this paper can be useful for the managers and planners who manage protected area–buffer zone resources.

This investigation has offered valuable information for the TDNP managers. The results can be used to prioritize land-use management projects funded by local and central governments and other non-governmental organizations. The study shows that GIS databases of different formats and sources can be integrated efficiently to establish a LSA for cropland. The methodology is useful for identifying priority areas for crop farming, and thus, it contributes to improving the efficiency of conservation and of sustainable land management. The approach can also be handy for land-use managers working in other protected areas in Vietnam that have similar conditions to the TDNP region. The land suitability information produced in recent research is valuable. However, land-use decisions are not only based on such information, but also on other assessments, such as an economic analysis of land uses and environmental impact assessments. Therefore, we recommend that future studies should consider these assessments in order to offer decision makers a comprehensive basis on which to orient a feasible strategy, and to make a sound decision towards a more sustainable TDNP.

References

Ahamed TRN, Rao KG, Murthy JSA (2000) GIS-based fuzzy membership model for cropland suitability analysis. Agric Syst 63:75–95

Bridle JR, Pedro PM, Butlin RK (2004) Habitat fragmentation and biodiversity: testing for the evolutionary effects of refugia. Evolution 58:1394–1396

TDMP (Tam Dao National Park and Buffer Zone Management Project) (2005) Rural household economics baseline survey 2005. Centre of Services, Eco-tourism and Environmental Education, Tam Dao. Retrieved on 2 March 2010, from http://tamdaonp.com.vn/

Colchester M (1997) Salvaging nature: indigenous peoples and protected areas. In: Ghimire KB, Pimbert MP (eds) Social change and conservation: environmental politics and impacts of national parks and protected areas. Earthscan, London, pp 97–130

Collins MG, Steiner FR, Rushmaan MJ (2001) Land-use suitability analysis in the United States: historical development and promising technological achievements. Environ Manag 28:611–621

Cover SJ (1991) Integrating multi-criteria evaluation with geographical information systems. Int J Geogr Inf Syst 5:321–339

Eastman JR (2009) IDRISI Taiga, guide to GIS and remote processing. Clark University, Worcester, 234–256

Eastman JR, Jin W, Kyem PAK, Toledano J (1992) Participatory procedures for multi-criteria evaluation in GIS. In: Proceedings of Chinese professionals in GIS 92, NCGIA, Buffalo, pp 281–288

Eastman JR, Jin W, Kyem PAK, Toledano J (1995) Raster procedures for multi criteria/multi objective decisions. Photogramm Eng Rem Sens 61:539–547

FAO (Food and Agriculture Organization of the United Nations) (1976) A framework for land evaluation. Soils Bulletin No. 32

FAO (Food and Agriculture Organization of the United Nations) (2005) Global forest resources assessment: progress toward sustainable forest management. Retrieved on 25 September 2008, from http://www.fao.org/DOCREP/008/a0400e/a0400e00.htm

Ghazoul J (1994) Frontier Vietnam environmental research report 1: Tam Dao Nature Reserve: results of a biological survey. Ministry of Agriculture and Rural Development, Hanoi

Groenemans R, Ranst EV, Kerre E (1997) Fuzzy relational calculus in land evaluation. Geoderma 77:283–298

ICEM (International Centre for Environmental Management) (2003) Vietnam national report on protected areas and development, pp 19–47

IUCN (International Union for the Conservation of Nature) (1994) Guidelines for protected area management categories

Jianbo L (2006) Energy balance and economic benefits of two agroforestry systems in northern and southern China. Agric Ecosyst Environ 116:255–262

Khang ND, Hoe H, Duc HD, Thin NN, Tien DD, Lanh VL, Huyen TH (2007) Tam Dao National Park. Agricultural Publishing House, Hanoi, pp 9–56 (in Vietnamese)

Liu YS, Wang JY, Guo LY (2006) GIS-based assessment of land suitability for optimal allocation in the Qinling Mountains, China. Pedosphere 16:579–586

Malczewski J (1999) Spatial multicriteria decision analysis. In: Spatial multicriteria decision making and analysis: a geographical information sciences approach. Ashgate Publishing, Aldershot, pp 102–269

Malczewski J (2006) GIS-based multicriteria decision analysis: a survey of the literature. Int J Geogr Inf Syst 20:703–726

McBratney AB, Odeh IOA (1997) Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85–113

McNeely JA (1993) Parks for life: report of the IVth world congress on national parks and protected areas

Meyfroidt P, Lambin FE (2008) The causes of the reforestation in Vietnam. Land Use Policy 25:182–197

Palma JHN, Graves AR, Burgess PJ, Keesman KJ, van Keulen H, Mayus M, Reisner Y, Herzog F (2007) Methodological approach for the assessment of environmental effects of agroforestry at the landscape scale. Ecol Eng 29:450–462

Pereira JMC, Duckstein LA (1993) Multiple criteria decision-making approach to GIS-based land suitability evaluation. Int J Geogr Inf Syst 7:407–424

Phuong NN, Dembner SA (1994) Improving the lifestyles of people in protected areas of Vietnam. Retrieved on 2 April 2010, from http://www.fao.org/docrep/v2900e/v2900e00.htm#Contents

Pimbert MP, Pretty JN (1997) Parks, peoples and professionals: putting participation into protected area management. In: Ghimire KB, Pimbert MP (eds) Social change and conservation: environmental politics and impacts of national parks and protected areas. Earthscan, London, pp 297–330

Quan B, Zhu HJ, Chen SL, Romkens MJM, Li BC (2007) Land suitability assessment and land use change in Fujian province, China. Pedosphere 17:493–504

Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York

Saaty TL (1990) The analytic hierarchy process: planning, priority setting, resource allocation. RWS Publications, Pittsburgh

Stainer F (1991) Landscape planning: a method applied to a growth management example. Environ Manag 15:519–529

Sui DZ (1993) Integrating neural networks with GIS for spatial decision making. Oper Geogr 11:13–20

van Ranst E, Tang H, Groenemans R, Sinthurahat S (1996) Application of fuzzy logic to land suitability for rubber production in peninsular Thailand. Geoderma 70:1–19

Wang X, Lu C, Fang J, Shen Y (2007) Implications for development of grain-for-green policy based on cropland suitability evaluation in desertification-affected north China. Land Use Policy 24:417–424

WCPA (World Commission on Protected Areas) (2010) Overview: what is a protected area? Retrieved on 3 April 2010, from http://www.iucn.org/about/union/commissions/wcpa/wcpaoverview/

Chapter 15
Neighborhood Interaction in Urban Land-Use Changes Using Cellular Automata-Based Geo-Simulation*

Yaolong Zhao, Bingliang Cui, and Yuji Murayama

15.1 Introduction

Cities can be understood as complex systems with intrinsic characteristics of emergence, self-organization, self-similarity, and non-linear behavior of land-use dynamics (Barredo et al. 2003; Batty 2005). Cities incessantly undergo a dynamic and complex process of urban land-use changes. This complex process has direct impacts on the urban environment (Jusuf et al. 2007; Pauleit et al. 2005), and may even profoundly disrupt the structure and function of ecosystems on a global scale (Lambin et al. 2001; Turner et al. 1990). Therefore, the complex spatial processes of urban land-use changes must be thoroughly understood in order to provide municipal and urban planners with a basis for assessing the ecological impacts of urban land-use changes, and to support spatial decision-making. For this purpose, various spatial dynamic models of urban land-use change, in particular cellular automata (CA), multi-agent systems (MAS), and geographical information system (GIS)-based urban geo-simulation models, have been constructed and successfully

*This chapter has been improved from "Yaolong Zhao, Bingliang Cui and Yuji Murayama (2011), Characteristics of neighborhood interaction in urban land-use changes: a comparative study between three metropolitan areas of Japan. *Journal of Geographical Sciences*, 21(1), 65–78. Copyright (2011), with permission from Springer."

Y. Zhao (✉) • B. Cui
School of Geography, South China Normal University,
Guangzhou, People's Republic of China
e-mail: yaolong@scnu.edu.cn; cui_bl@163.com

Y. Murayama
Division of Spatial Information Science, Graduate School of Life
and Environmental Sciences, University of Tsukuba,
Tsukuba, Ibaraki, Japan
e-mail: mura@geoenv.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_15,

applied to many cities (Barredo and Demicheli 2003; Batty et al. 1999; Torrens 2006; White and Engelen 2000; Yeh and Li 2002). In such geo-simulation models, neighborhood interaction is an important component (Batty 1991; Wu 1998; Zhao and Murayama 2007). Neighborhood interaction means local spatial interactions between neighborhood land-use categories such as facilities, residential areas, and industries in urban areas. Here, "neighborhood" means "close to", i.e., neighborhood land-use parcels may or may not be contiguous (touching). Such interaction has a great impact on the spatial processes of urban land-use changes (Batty 2005; Couclelis 1989). This type of factor is known as the neighborhood effect of urban land-use changes. The neighborhood effect plus exogenous factors (like spatial interactions between cities) and endogenous factors (like transportation networks in urban areas) determine the spatial process of urban land-use change (White and Engelen 2000). Furthermore, it is often cited as the main factor which decides urban land-use change patterns, since other factors are comparatively stable in the spatial process of urban land-use change during a set period.

In fact, neighborhood interactions are always a focus in the field of CA research. This standpoint can be derived from the basic definition of CA offered in von Neumann's lecture of 1951 (von Neumann 1951). There are four elements in a basic CA structure, namely, automata size, state, neighborhood, and the transition rule of automata state. Under a certain size and prescribed state, automata dynamics is controlled by the transition rule, which is established only by considering the interaction of automata in the neighborhood area. In the early phase of the application of CA to urban studies, urban geo-simulation models came from the basic definition of CA (Batty 1991; Couclelis 1989; Phipps 1989). At that time, urban geo-simulation models were mainly used to explore the intrinsic characteristics of urban systems such as self-organization and self-similarity (Batty and Longley 1994; Batty and Xie 1994; White and Engelen 1993, 1994). Later, scholars moved to focus on the simulation of actual urban land-use dynamics using CA models, and therefore CA models were updated by adding components or adjusting neighborhood configurations for this purpose (Batty et al. 1999; Clarke et al. 1997; White and Engelen 2000; Wu 1998). Some scholars also proposed MAS for urban land-use geo-simulation to overcome the weakness of the CA approach (Le et al. 2008; Parker et al. 2001). However, neighborhood interaction still is deemed an important component in urban geo-simulation models whether they are based on MAS or updated CA (Torrens and Benenson 2005).

Although neighborhood interaction has been highlighted in studies of urban land-use changes (White and Engelen 2000; Yang and Billings 2000; Zhao and Murayama 2007), there is still much that is unknown about its characteristics, although these provide the basic information which constitute the neighborhood effect rules for urban geo-simulation models. For example, neighborhoods in urban geo-simulation models generally adopt either the von Neumann 3×3 (or 5×5) or the Moore 3×3 configuration for simplicity (Batty 1998; Wu 1998; Yeh and Li 2001). Some scholars have enlarged the size of the configuration in neighborhood interaction models to give sufficient consideration to the human characteristics of urban systems (White and Engelen 1993; Zhao and Murayama 2007). However, few

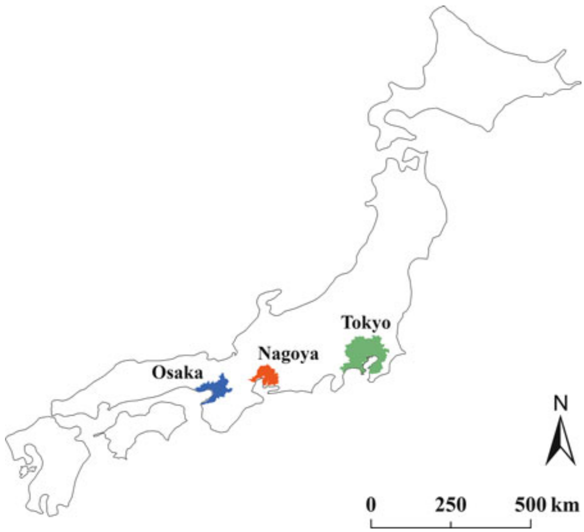


Fig. 15.1 Locations of the three Japanese metropolitan areas

studies have focused on the reason why such neighborhood configurations are selected and modeled, and what the mechanism of the neighborhood effect is. In particular, there are very few discussions in the literature about whether the issues that are characteristic of neighborhood interactions are the same in different cities. The answer to this question is very important for an understanding of the mechanism of the neighborhood effect on urban land-use changes, and for constructing a universal urban geo-simulation model which may be applied to any city in Japan. This research focuses on this issue, and tries to interpret the similarities and differences in the characteristics of neighborhood interactions in urban land-use changes by comparing three metropolitan areas in Japan, i.e., Tokyo, Nagoya, and Osaka, using such aids as the neighborhood interaction model and the similarity measure function. The results of this research will provide important information for constructing effective and operational neighborhood effect modules in urban geo-simulation systems.

15.2 Methodology

15.2.1 Study Areas

The three Japanese metropolitan areas of Tokyo, Nagoya, and Osaka were selected for a comparative study. Figure 15.1 shows the location and range of the study areas. These three metropolitan areas are the business, economic, political, and population centers of Japan. In the period after World War II, in particular, a large proportion of the Japanese population congregated in these areas (Murayama 2000), so that by 2005, these three areas accounted for more than 50% of Japan's total population.

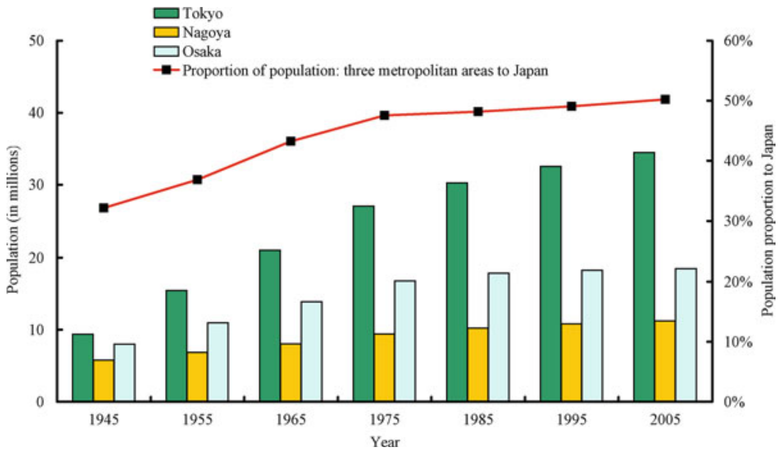


Fig. 15.2 Population increase in the three metropolitan areas and their proportions of Japan's population from 1945 to 2005 (*Source*: Statistics Bureau, Ministry of Internal Affairs and Communications of Japan)

Figure 15.2 illustrates the population increase in the three metropolitan areas and their respective proportions of the population of the whole country from 1945 to 2005. With the increases in population, the urbanized areas expanded into the surrounding regions of the three metropolitan areas at an astounding pace. As well as urban growth, urban functions in the existing urban areas also experienced a process of self-adjustment (Takahashi and Taniuchi 1994). Therefore, these three metropolitan areas are appropriate study areas for gaining an understanding of the spatial processes of urban land-use changes in Japan, as well as the characteristics of neighborhood interactions at that time.

15.2.2 Data Set

Detailed digital information of the metropolitan areas (10 m grid land-use; DDIMA10 m) of Tokyo, Nagoya, and Osaka was produced by the Geographical Survey Institute of the Ministry of Construction of Japan. DDIMA10 m of Tokyo was investigated in 1974, 1979, 1984, 1989, and 1994; DDIMA10 m of Nagoya in 1977, 1982, 1987, 1991, and 1997; and DDIMA10 m of Osaka was investigated in 1974, 1979, 1985, 1991, and 1996. DDIMA10 m provides an abundant and detailed urban land-use classification system, which includes a range of socio-economic information over a period of time. There are 15 categories of land-use in these data sets, namely (A) forest and wasteland, (B) paddy field, (C) dry field and other farmland, (D) construction areas, (E) vacant land, (F) industrial land, (G) low-storey residential land, (H) densely developed low-storey residential land, (I) medium and high-storey residential land, (J) commercial land, (K) roads, (L) parks, (M) public facilities, (N) water, and (O) other. Here, in order to reach a clear understanding of the main characteristics of urban land-use changes, land-use is grouped into the following ten categories: (1) vacant land, (2) industrial land, (3) residential land, (4) commercial land, (5) roads, (6) public land, (7) special land, (8) forest and wasteland, (9) cropland, and (10) water. Here, land-uses (D) and (E) in the original data set are combined into land-use (1); (G), (H), and (I) into (3); (L) and (M) into (6); and (B) and (C) into (9). The others remain in their original categories.

The grouped land-use classification system reflects the intrinsic characteristics of urban land-use changes. "Water" represents fixed features, i.e., it is assumed that it will not change, and therefore it is not involved in land-use dynamics in order to protect the living environment. Forest, wasteland and cropland are passive features that play a role in the process of land-use changes, but the changes are not driven by an exogenous demand for land. They appear or disappear in response to active functions of land being used or abandoned. The active functions are four land-use categories which are forced into existence by demands for land generated exogenously in response to changes in the urbanized areas: vacant, industrial, residential, and commercial land. Roads, public, and special lands are active features which are the dynamics of the model, but they are mainly controlled by the municipal government through urban land-use planning.

15.2.3 Data Processing

Urban land-use changes are driven by multiple factors such as urban and region planning policy, environmental characteristics, local-scale neighborhood characteristics, the spatial characteristics of cities, and so forth (Carver 1991; Voogd 1983). These factors can be divided into two types: natural forces and human activities. Over a short time-scale, the effects of both natural forces and human activities are comparatively stable for a certain area (Zhao et al. 2010). Therefore, short time-intervals, here about 5 years, are used to extract land-use change patterns in order to interpret the characteristics of neighborhood interactions. The latest time-intervals are selected from DDIMA10 m data sets as follows: Tokyo Metropolitan Area, 1989–1994; Nagoya Metropolitan Area, 1991–1997; Osaka Metropolitan Area, 1991–1996. Although there are slight differences in the study periods in these metropolitan areas, land-use policies in the whole country during such short periods do not change very much. Therefore, it is assumed that the tiny differences will not influence the understanding of the characteristics of neighborhood interactions.

Considering the huge data sets from the three metropolitan areas and the resulting time-consuming computations, the grid size of the land-use data set was set at 100 m × 100 m by aggregating the original 10 m × 10 m cells by a majority rule (in the process of aggregation, the land-use category of a 100 m × 100 m grid area is determined by the maximum proportion of land-use categories of the 10 m × 10 m grids which are located in that 100 m × 100 m grid area).

Urban land-use changes in the three metropolitan areas were extracted from data sets from two adjacent time-sections. Land-use changes in the area at a distance of less than 600 m to the boundary of the study area were deleted in order to eliminate the boundary effect. Appropriate sample numbers were selected randomly in order to create an approximately 1:1 ratio of transformed to non-transformed cells.

15.2.4 Neighborhood Interaction Model

"Neighborhood" has no determinate configuration in many correlative studies (Barredo et al. 2003; Batty 1998; Li and Yeh 2001; White and Engelen 1997). According to Zhao and Murayama (2007), an extended neighborhood configuration is defined as an area within a radius of eight cells from the central developable cell in a model which contains 196 cells. The contribution of one cell in the neighborhood is associated with its state and its distance to the central developable cell *i* based on Tobler's first law of geography (Tobler 1970). The neighborhood effect on the probability *N* of the conversion of a cell to land-use *k* is described as a function of a set of aggregated effects of cells in the neighborhood:

$$\mathrm{Log}\left(\frac{N_{ik}}{1-N_{ik}}\right)=\beta'_{0i}+\sum_{k}\beta'_{ihk}\sum_{m}\frac{A_m}{d_{mi}^2}I_{mh} \tag{15.1}$$

where m is the number of cells in the neighborhood, A_m is the area of cell m (here in square meters), d_{mi} is the Euclidean distance between the central developable cell i and cell m in the neighborhood area, β'_{ihk} is the constant of the effect of land-use h on the transition to land-use k, + stands for positive, – stands for repulsive, I_{mh} is the index of cells, and $I_{mh}=1$ if the state of cell m is equal to h; otherwise $I_{mh}=0$. β'_{0i} and β'_{ihk} are the coefficients to be calibrated with a maximum likelihood estimation. The coefficients stand for the effects of different land-use categories in the neighborhood on the change in transformation odds ($N_{ik}/(1-N_{ik})$) of central cell i to land-use category k. If β'_{ikh} is positive, the odds will add to the increase in the aggregated effect of land-use type k, and vice versa. If one of the coefficients does not pass the hypothesis test at the 0.05 level, $\beta'_{ikh}=0$, indicating that the corresponding land-use category does not affect the transformation of the central developable cell i. The values of the coefficients represent the intensity of the effect on the transformation odds. The greater the value, the more intense the effect. Obviously, the coefficients are suitable indices which can be used to analyze the effect of land-use categories in neighborhoods on the transformation of cells. Herein, the coefficients are used to interpret the neighborhood interactions in urban land-use changes in the three metropolitan areas.

15.2.5 Similarity Measure Function

As this research focuses on land-use changes in urbanized areas, all seven urban land-use categories, i.e., vacant, industrial, residential, commercial, road, public, and special land, in neighborhood areas should affect the transformation of the four active land-use categories in metropolitan areas. In these seven urban land-use categories, special land comprises military, royal, and other special land which is always closed to the public. Therefore, it is assumed that the effect of special land on the transformation of active land-use categories is very limited and can be omitted from an understanding of urban land-use changes. Accordingly, the values of the six remaining effect coefficients β'_{ihk} can be obtained for one active land-use category in any metropolitan area.

The values of the coefficients for different metropolitan areas are compared using the high-dimension similarity measure function Hsim(*X*, *Y*) (Yang and Zhu 2004):

$$\mathrm{Hsim}(X,Y) = \frac{\sum_{i=1}^{d} \frac{1}{1+|x_i - y_i|}}{d} \tag{15.2}$$

where X and Y are two sets (objects) with dimension d which are compared for similarity, and x_i and y_i stand for the data of X and Y in the ith dimension.

This function represents the degree of similarity of two objects X and Y. The higher the value of Hsim(*X*, *Y*), the more similar the two objects are. If the minimum value of Hsim(*X*, *Y*) is 0, X and Y are not similar at all. The maximum value of Hsim(*X*, *Y*) is 1, meaning that X and Y are identical. In this research, X and Y stand for the same active land-use category in different metropolitan areas, and x_i or y_i stand for the values of the coefficients of the neighborhood interaction in urban land-use changes. Higher values of Hsim(*X*, *Y*) indicate a higher degree of similarity of neighborhood interactions in different metropolitan areas.

15.3 Results and Discussion

15.3.1 Urban Land-Use Structure and Changes in the Three Metropolitan Areas

Land-use patterns and structure in the three metropolitan areas showed a similar mode in the base years of 1989 in Tokyo, 1991 in Nagoya, and 1991 in Osaka (Fig. 15.3a, c, e). Residential land is dominant in the urban land-use structure of these three metropolitan areas, accounting for more than 43% of urban land in the three areas. The area proportion of residential land in Tokyo even reached 48.7%. The high values of this proportion illustrate the residential function of the metropolitan areas in Japan. The area of public land is the second highest proportion at more than 14%. Public land in urbanized areas mainly includes public service facilities (such as educational facilities, city hall) and open spaces like parks. The higher proportion of public land in the metropolitan areas indicates the efforts of municipal governments to provide residents with more public service facilities and open spaces.

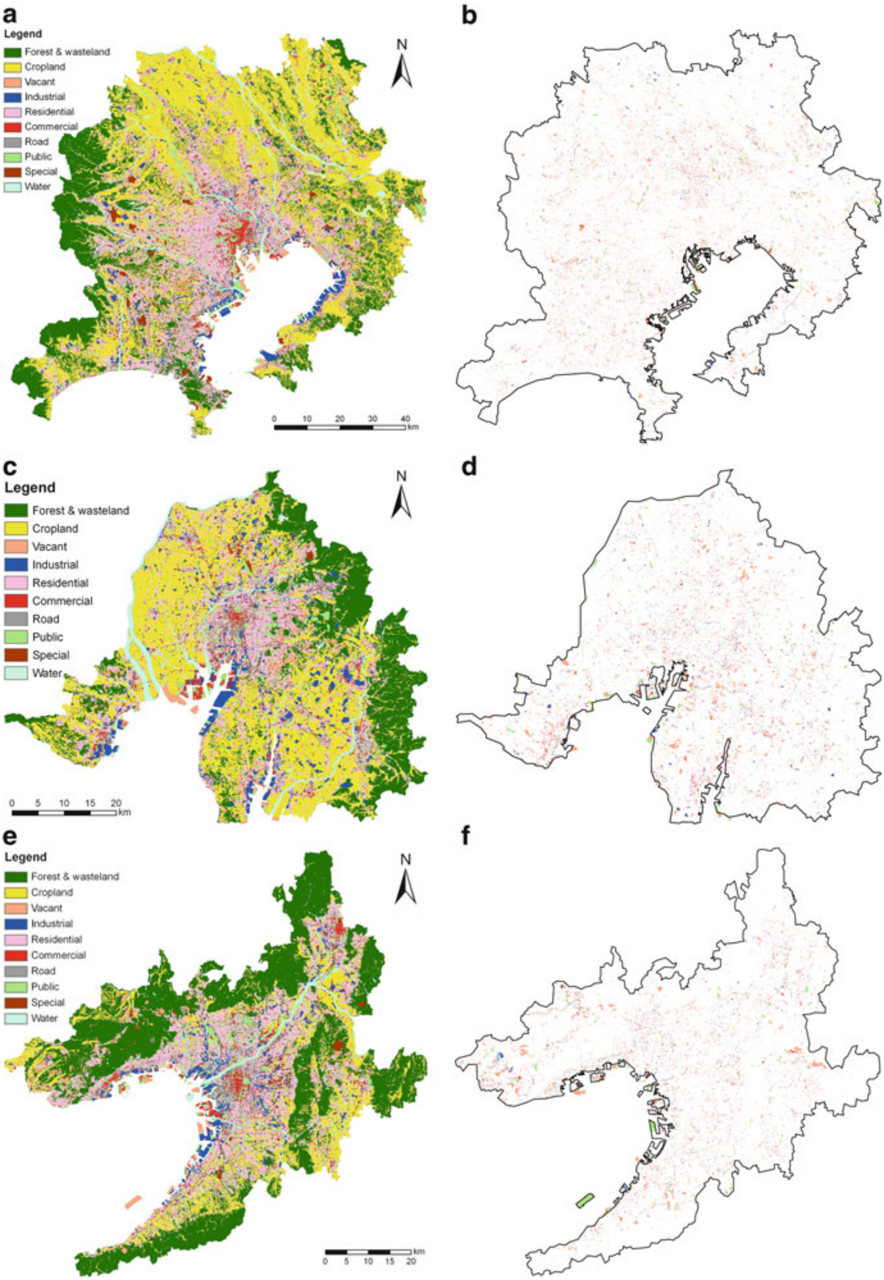
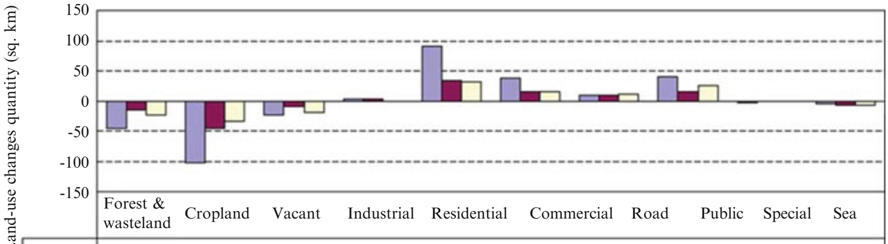


Fig. 15.3 Land-use patterns in the three metropolitan areas in the base year land-use map of: (**a**) Tokyo 1989; (**c**) Nagoya 1991; (**e**) Osaka 1991; and land-use changes (*changed areas*) in: (**b**) Tokyo from 1989 to 1994; (**d**) Nagoya from 1991 to 1997; (**f**) Osaka from 1991 to 1996



	Forest & wasteland	Cropland	Vacant	Industrial	Residential	Commercial	Road	Public	Special	Sea
Tokyo	-46.87	-102.85	-24.03	4.00	90.93	37.02	9.22	40.19	-2.84	-4.80
Nagoya	-14.70	-45.07	-9.45	3.52	34.20	14.79	8.32	15.44	-0.01	-7.07
Osaka	-23.70	-32.73	-18.50	-1.02	30.65	15.16	11.23	25.14	-0.01	-6.30

Land-use category and changes quantity (sq. km)

Fig. 15.4 Urban land-use changes in the three metropolitan areas (Tokyo, 1989–1994; Nagoya, 1991–1997; Osaka, 1991–1996)

The third highest proportion is vacant land, which represents land being prepared for, or subjected to, construction. Its proportion in urbanized areas is generally more than 10%, indicating the potential dynamics of urban areas in the three metropolitan areas. Industrial land is always either located along the coast of the three metropolitan areas or agglomerated in the suburbs. The proportion of industrial land to urbanized areas in Tokyo and Osaka is less than 10%, whereas that in Nagoya is at a higher level of 13.6%. Compared with other metropolitan areas, Nagoya is a city with an agglomeration of industries, especially automobile industries. The higher area proportion of industrial land indicates the degree of industrialization in Nagoya.

Commercial land is generally located in the center of metropolitan areas and near subway or railway stations. The proportion of commercial land to urbanized area in one metropolitan area is about 7%. The CBD (central business district) in the center of the metropolitan area and the sub-CBD near main subway stations are notable in land-use maps of the three metropolitan areas (Fig. 15.3a, c, e). The proportion of road land in the three metropolitan areas is not low, reflecting the high density of roads in metropolitan areas in Japan. The proportion of special land is not significant except for Tokyo with 1.7%.

The spread of urbanized areas has different development potential in the three metropolitan areas in terms of the whole land-use patterns (Fig. 15.3a, c, e). Agricultural land accounts for the main proportion of suburban areas in Tokyo and Nagoya. These areas are flat with smooth topography, which provides prime development space. However, the suburban areas in Osaka are mainly mountainous areas where urban land-use patterns are limited by the rugged topography. Urban areas cannot easily spread to such places.

The changes in land-use structure and patterns mainly show the trend in urban growth during the study period in the three metropolitan areas (Figs. 15.3b, d, f, and 15.4). Areas of cropland, vacant land, forest, and wasteland decreased to 173.75 km^2 in Tokyo, to 69.22 km^2 in Nagoya, and to 74.93 km^2 in Osaka. Increases in residential, commercial, and public land are notable. Most of the land parcels of cropland, forest, and wasteland which had decreased had become residential and public land. The residential and public-oriented strategy of urban development is well illustrated by this phenomenon. In addition, as the three metropolitan areas are coastal cities, the urbanized areas also spread out into the sea. During the relatively short study period, a considerable amount of the marine area was reclaimed and converted into urban area: 4.80 km^2 in Tokyo, 7.07 km^2 in Nagoya, and 6.30 km^2 in Osaka (Fig. 15.4). Except for industrial land in Osaka, all of the urbanized areas of industrial, residential, commercial, roads, and public land in the three metropolitan areas increased, but at different rates. Land covered by roads only increased a little during the 5 years, and industrial land showed less change than residential or commercial land in the three metropolitan areas. Industrial land in Osaka decreased slightly during the period. Patterns of land-use changes showed varying characteristics. The land-use parcels which changed were widely dispersed, and did not agglomerate in the study areas (Fig. 15.3b, d, f).

15.3.2 Characteristics of Neighborhood Interaction in Urban Land-Use Changes

The coefficient β'_{ikh} of different land-use categories in the neighborhoods was regressed at the 0.05 level of the hypothesis test for urban land-use changes in the four active categories in the three metropolitan areas. The coefficient β'_{ikh} stands for the intensity of the neighborhood effect on the probability of an active land-use category, i.e., the neighborhood interaction in urban land-use changes. Considering its particular nature, special land is assumed not to be involved in the interaction. Figure 15.5 shows the results of regression. Here, the horizontal axis stands for the land-use categories which affect the transformation of the four active land-use categories in the neighborhood. The vertical axis represents the value of the regression coefficient β'_{ikh}. Figure 15.5 shows that the value of the regressed coefficient of each active land-use category on its own transformation is always more than that of the other land-use categories, especially industrial and commercial land. For vacant land, the values of the coefficient itself in the three metropolitan areas are more than 0.5, whereas those of other land-use categories are less than 0.4. This difference is similar for residential land, but is more notable for industrial land. For industrial land, the values of the coefficient itself are near to, or greater than, 1.0, while those of other categories are less than 0.5.

The values of the coefficient of commercial land also are greater than those of other land-use categories for commercial land. This phenomenon represents the effect of spatial autocorrelation in the spatial process of urban land-use changes in the three metropolitan areas (Zhao and Murayama 2006), which cause the spatial aggregation of the same urban land-use category (Herold et al. 2005; Palivos and Wang 1996). This result is in line with the characteristics of agglomeration of industrial and commercial land allocation in Japan (Baba and Shibuya 2000; Ida 2006), and also certifies the effectiveness of the neighborhood interaction model selected.

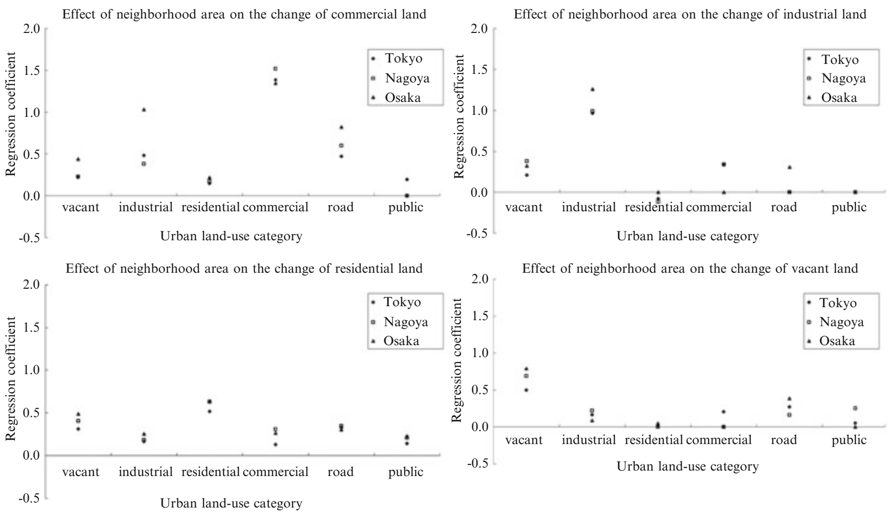


Fig. 15.5 Value of the regression coefficient of neighborhood interactions for the three metropolitan areas

The characteristics of the neighborhood effect on changes in urban land-use categories for different active land-use categories are quite different among the metropolitan areas, meaning that the mechanism of urban land-use changes is different from that of land-use categories. For changes in vacant land, the intensity of the effect of land-use categories in the neighborhood is less strong. The effect of other land-use categories in the neighborhood is close to 0. This shows that in the change process, vacant land rarely interacts with other land-use categories in the same neighborhood area. Moreover, the values of the percentage correctly predicted (PCP) and the relative operating characteristic (ROC) of logistical regression tests for vacant land are smaller than those for other land-use categories. This may be because of the complex definition of vacant land, and it also indicates that the location of vacant land would mainly be determined by factors other than neighborhood effects.

Industrial land shows similar characteristics to vacant land. However, the intensity of the effect of industrial land in the neighborhood on itself is stronger than the effect of vacant land on itself, indicating the strong degree of spatial aggregation of industrial land. The effects of other land-use categories in the neighborhood on residential land show approximately the same intensity, whereas the effect on commercial land shows a different intensity. Roads and industrial and commercial land in the neighborhood obviously impact on the allocation of commercial land.

The differences in neighborhood interactions among urban land-use categories suggest that attention should be paid to the land-use classification system when constructing an urban geo-simulation model. An urban area is composed of many categories of land-use. Owing to the difficulty of obtaining land-use information in urbanized areas or other reasons (Batty 1971), urbanized areas are classified as one

Table 15.1 Comparison of the similarity measure of neighborhood interaction in urban land-use changes between the three metropolitan areas during the study period

Type of land	Tokyo–Nagoya	Nagoya–Osaka	Tokyo–Osaka
Vacant	0.889	0.894	0.892
Industrial	0.965	0.857	0.852
Residential	0.926	0.958	0.906
Commercial	0.914	0.843	0.823

Table 15.2 Comparison of the similarity measure of neighborhood interaction in urban land-use changes between the three metropolitan areas during the late 1980s

Type of land	Tokyo–Nagoya	Nagoya–Osaka	Tokyo–Osaka
Vacant	0.904	0.886	0.924
Industrial	0.826	0.785	0.923
Residential	0.911	0.925	0.921
Commercial	0.950	0.897	0.882

land-use category in much of the literature (Benenson 2007; Clarke et al. 1997). Thus, the mechanism of interaction among different urban land-use categories may be concealed. This makes it more difficult to understand the real processes and mechanisms of urban land-use changes, and may even lead to mistakes. Zhao and Murayama (2006) analyzed the characteristics of the effects of land-use classification systems on spatial patterns of land use.

Figure 15.5 illustrates that neighborhood interaction for one active land-use category in the spatial process of urban land-use changes generally shows similar characteristics in the three metropolitan areas, although neighborhood interaction is different for every other active land-use category. The similarity of neighborhood interaction in urban land-use changes between the three metropolitan areas during the study period was calculated using the similarity measure function Hsim(X, Y) to quantitatively describe the degree of similarity (Table 15.1). In the similarity measure function Hsim(X, Y), the values of coefficient β'_{ikh} of land-use categories of vacant, industrial, residential, commercial, roads, and public land comprise a vector with six dimensions. All the values of similarity measures between metropolitan areas were at least 0.823, with one even reaching 0.965, indicating a high degree of similarity of neighborhood interaction in urban land-use changes between the three metropolitan areas, although their land-use change patterns were not correlated.

The similarity measure of neighborhood interaction in urban land-use changes between the three metropolitan areas during the late 1980s (Tokyo, 1984–1989; Nagoya, 1987–1991; Osaka, 1985–1991) was calculated to avoid error and contingency in the calculation of the similarity measure function above (Table 15.2). The similarity measure during the period also shows high values of at least 0.826, except for 0.785 for the industrial land between the Nagoya and Osaka metropolitan areas. This result validates the finding that the characteristics of neighborhood interaction in urban land-use changes are similar in different urban areas. It may be concluded that the finding is universal for all cities in Japan.

This finding may have two meanings in the study of urban land-use changes in Japan. One is that the mechanism of urban land-use changes in neighborhood interaction at the local level would be similar for different cities. The other is that in the creation of urban geo-simulation systems, the values of the coefficients of neighborhood effect in urban land-use changes may be applied to different cities in Japan.

15.4 Concluding Remarks

Tokyo, Nagoya, and Osaka are three main metropolitan areas in Japan. Land-use patterns and structure show similar characteristics in the three metropolitan areas. Residential and public land occupies about 60% of the acreage of each metropolitan area. Moreover, in the process of urban growth, most of the lost agriculture and forest land was also transformed into residential and public use. The function of human habitation and the principle of priority of public use in the cities of Japan are empirically demonstrated by the proportion and trend of land-use change.

A city is a huge, complex system. One of the complexities is that urban areas are composed of different kinds of urban land-use categories. The dynamics of these categories form the whole pattern of urban land-use changes. Therefore, an investigation of the mechanism of urban land-use changes should start by exploring the dynamics of every urban land-use category. In this research, the characteristics of the neighborhood effect of urban land-use categories on the changes to four active land-use categories were quite different in each of the three metropolitan areas, illustrating the differences in the mechanisms of the changes in urban land-use categories in the cities. Consequently, the land-use classification system plays an important role in providing a clear understanding of the real mechanism of urban land-use changes. Suitable classification systems should be examined when considering urban land-use changes.

The spatial aggregation of urban land-use categories is illustrated by the spatial autocorrelation of the land-use category presented in the characteristics of neighborhood interaction. Industrial and commercial land displays a particularly strong aggregation of land-use allocation compared with other urban land-use categories in all three metropolitan areas. This finding agrees with the characteristics of agglomeration of industrial and commercial land allocation in Japan, and validates the effectiveness of the neighborhood interaction model.

In the three metropolitan areas, urban land-use changes have different spatial patterns, and urban land-use change patterns do not correlate with each other. However, neighborhood interaction in urban land-use changes generally shows similar characteristics for the transformation of every active land-use category. This implies that neighborhood interaction in any of these three metropolitan areas may be used to understand the mechanism of urban land-use changes in other cities of Japan. These results provide very useful material for constructing universal urban geo-simulation models which may be applied to any city in Japan. Nevertheless, the reason(s) why neighborhood interaction in urban land-use changes in the three metropolitan areas showed similar characteristics need(s) to be further investigated.

References

Baba Y, Shibuya M (2000) Tokyo game-soft cluster: analysis on firm's spatial agglomeration. Jpn Soc Sci Policy Res Manag 14(4):266–278 (in Japanese)
Barredo JI, Demicheli L (2003) Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos. Cities 20(5):297–310
Barredo JI, Kasanko M, McCormick N et al (2003) Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. Landsc Urban Plann 64(3):145–160
Batty M (1971) Modeling cities as dynamic systems. Nature 231:426–428
Batty M (1991) Cities as fractals: simulating growth and form. In: Crilly T, Earnshaw RA, Jones H (eds) Fractals and chaos. Springer, New York, pp 41–69
Batty M (1998) Urban evolution on the desktop: simulation with the use of extended cellular automata. Environ Plann A 30:1943–1967
Batty M (2005) Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. MIT, Cambridge
Batty M, Longley PA (1994) Fractal cities: a geometry of form and function. Academic, London
Batty M, Xie Y (1994) From cells to cities. Environ Plann B 21:s31–s48
Batty M, Xie Y, Sun Z (1999) Modeling urban dynamics through GIS-based cellular automata. Comput Environ Urban Syst 23(3):205–233
Benenson I (2007) Warning! The scale of land-use CA is changing! Comput Environ Urban Syst 31(2):107–113
Carver SJ (1991) Integrating multi-criteria evaluation with geographical information systems. Int J Geogr Inform Syst 5:321–339
Clarke KC, Hoppen S, Gaydos L (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environ Plann B 24:247–261
Couclelis H (1989) Macrostructure and microbehavior in a metropolitan area. Environ Plann B 16:141–154
Herold M, Couclelis H, Clarke KC (2005) The role of spatial metrics in the analysis and modeling of urban land use change. Comput Environ Urban Syst 29(4):369–399
Ida N (2006) A spatial econometric analysis of manufacturing agglomeration in Osaka. Doshisha Univ Econ Rev 57(3):295–318 (in Japanese)
Jusuf SK, Wong NH, Hagen E et al (2007) The influence of land use on the urban heat island in Singapore. Habitat Int 31(2007):232–242
Lambin EF, Turner BL, Geist HJ et al (2001) The causes of land-use and land-cover change: moving beyond the myths. Glob Environ Chang 11(4):261–269
Le QB, Park SJ, Vlek PLG et al (2008) Land-use dynamic simulator (LUDAS): a multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. I. Structure and theoretical specification. Ecol Inform 3(2):135–153
Li X, Yeh AGO (2001) Calibration of cellular automata by using neural networks for the simulation of complex urban systems. Environ Plann A 33:1445–1462
Murayama Y (2000) Japanese urban system. Kluwer Academic, Dordrecht
Palivos T, Wang P (1996) Spatial agglomeration and endogenous growth. Reg Sci Urban Econ 26(6):645–669
Parker DC, Evans TP, Meretsky V (2001) Measuring emergent properties of agent-based landuse/landcover models using spatial metrics. In: Seventh annual conference of the International Society for Computational Economics, Yale University
Pauleit S, Ennos R, Golding Y (2005) Modeling the environmental impacts of urban land use and land cover change: a study in Merseyside, UK. Landsc Urban Plann 71(2–4):295–310
Phipps M (1989) Dynamic behavior of cellular automata under the constraint of neighborhood coherence. Geogr Anal 21:197–215
Takahashi N, Taniuchi T (1994) The three metropolitan areas in Japan: changing spatial structures and future perspectives. Kokon Syoin, Tokyo, in Japanese
Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Geogr Anal 46(2):234–240

Torrens PM (2006) Simulating sprawl. Ann Assoc Am Geogr 96(2):248–275
Torrens PM, Benenson I (2005) Geographic automata systems. Int J Geogr Inform Sci 19(4):385–412
Turner BL II, Kasperson RE, Meyer WB et al (1990) Two types of global environmental change: definitional and spatial-scale issues in their human dimensions. Glob Environ Chang 1(1):14–22
von Neumann J (1951) The general and logical theory of automata. In: Jeffress LA (ed) Cerebral mechanisms in behavior: the Hixon symposium. Wiley, New York, pp 1–41
Voogd H (1983) Multicriteria evaluation for urban and regional planning. Pion, London
White R, Engelen G (1993) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environ Plann A 25:1175–1199
White R, Engelen G (1994) Urban systems dynamics and cellular automata: fractal structures between order and chaos. Chaos Solitons Fract 4(4):563–583
White R, Engelen G (1997) Cellular automata as the basis of integrated dynamic regional modeling. Environ Plann B 24:235–246
White R, Engelen G (2000) High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Comput Environ Urban Syst 24(5):383–400
Wu F (1998) SimLand: a prototype to simulate land conversion through integrated GIS and CA with AHP-derived transition rules. Int J Geogr Inform Sci 12(1):63–82
Yang Y, Billings SA (2000) Neighborhood detection and rule selection from cellular automata pattern. IEEE Trans Syst Man Cybern Syst Hum 30:840–847
Yang F, Zhu Y (2004) An efficient method for a similarity search on quantitative transaction data. J Comput Res Dev 41(2):361–368 (in Chinese)
Yeh AGO, Li X (2001) A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. Environ Plann B 28:733–753
Yeh AGO, Li X (2002) A cellular automata model to simulate development density for urban planning. Environ Plann B 29:431–450
Zhao Y, Murayama Y (2006) Effect of spatial scale on urban land-use pattern analysis in different classification systems: an empirical study in the CBD of Tokyo. Theory Appl GIS 14(1):29–42
Zhao Y, Murayama Y (2007) A new method to model neighborhood interaction in cellular automata-based urban geosimulation. Lect Notes Comput Sci 4488:550–557
Zhao Y, Dong F, Zhang H (2010) Should the neighborhood effect be stable in a urban geosimulation model? A case study of Tokyo. Lect Notes Comput Sci 6016:134–143

Chapter 16
Web-Based Interactive Walkability Measurement Using Remote Sensing and Geographical Information Systems*

Ko Ko Lwin and Yuji Murayama

16.1 Introduction

The concept of walkability conveys how conducive the built environment is to walking. It has been adopted in many parts of the world to predict people's physical activity and mode of transportation (Frank and Engelke 2005; Owen et al. 2004; Sallis et al. 2004). Walkability captures the proximity between functionally complementary land uses (live, work, and play) and the directness of a route or the connectivity between destinations (Forsyth and Southworth 2008; Moudon et al. 2006). A walk score is an indicator of how "friendly" an area is for walking. This score is related to the benefits to society in terms of energy savings and improvements in health that a particular environment offers to its residents. For example, a recently developed walk score web site uses Google Maps, specifically Google's local search application programming interface (API), to find stores, restaurants, bars, parks, and other amenities within walking distance of any address entered. The walk score currently includes addresses in the United States, Canada, and the United Kingdom. The algorithm behind this score indicates the walkability of a given route based on the fixed distance from one's home to nearby amenities. The number of amenities found nearby is the leading predictor of whether people will walk rather than take another travel mode. However, evaluating walkability is challenging because it requires the consideration of many subjective factors (Reid 2008). Moreover, all technical disciplines related to walkability have their own terminology and jargon (Abley 2005).

*This chapter has been improved from "Ko Ko Lwin and Yuji Murayama (2011), Modelling of urban green space walkability: eco-friendly walk score calculator. *Computers, Environment and Urban Systems*, 35(5), 408–420." Copyright (2011), with permission from Elsevier.

K.K. Lwin (✉) • Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kokolwin2002@yahoo.com; mura@geoenv.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_16, © Springer Japan 2012

Neighborhood environmental quality is an important factor affecting human health. Fortunately, neighborhood environmental quality can be improved by proper urban management. Thus, epidemiological studies have explored the relationship between access to nature and health. For example, a study in Sweden by Grahn and Stigsdotter (2003) demonstrated that the more often one visits green areas, the less often one reports stress-related illness. An epidemiological study performed in The Netherlands (Maas et al. 2006) showed that residents of neighborhoods with abundant green spaces tended, on average, to enjoy better general health. Another possible mechanism relating nature to health occurs during social interactions and social cohesion. Several studies conducted in Chicago suggest that green spaces, especially trees, may facilitate positive social interactions between neighborhood residents (Kweon et al. 1998).

Therefore, in many parts of the world, current urban planning activities are shifting toward a focus on "green" living. Many cities around the world are now developing integrated solutions to major environmental challenges, and are transforming themselves into more sustainable and self-sufficient communities (Dizdaroglu et al. 2009). Among the environmental benefits achieved by such green interventions are the following: reduced cooling and heating demand, improved air quality, reduced storm-water runoff, the enrichment of urban biodiversity and urban agriculture, a reduced urban heat-island effect, a contribution to carbon-neutral architecture, an aesthetic improvement to the skylines of cities, and the economic impact of the green spaces (Roehr and Laurenz 2008).

Given the great interest in walking activities and other urban sustainability measures, the purpose of this chapter is to develop an integrated methodology [using remote sensing, geographical information systems (GIS), and spatial web technology] to model urban green space walkability, which enables local residents to make informed decisions that will improve their living conditions and physical health. The proposed methodology uses advanced land observing satellite (ALOS) data to identify the green spaces, which are then integrated with other GIS data sets, such as road networks, public facility locations, and building footprints, to calculate an eco-friendly walk score to enable residents to make decisions using an interactive web-based GIS. We use Tsukuba City, Japan, as a case study.

This chapter is organized into three sections. Following the introduction, the conceptual framework is presented. This is followed by a discussion of the case study, including a list of data used, data processing steps, and a description of the implementation of the system. We proceed to explain the GIS analytical functions of our eco-friendly walk score calculator and its potential applications. The results of a qualitative usability study are presented in Sect. 16.5, while our conclusions are discussed in Sect. 16.6.

16.2 Conceptual Framework

Recently, GIS studies of urban green space areas have been increasing in number. For example, Mahon and Miller (2003) used GIS to identify green space areas with high ecological, recreational, and aesthetic values to protect certain green space

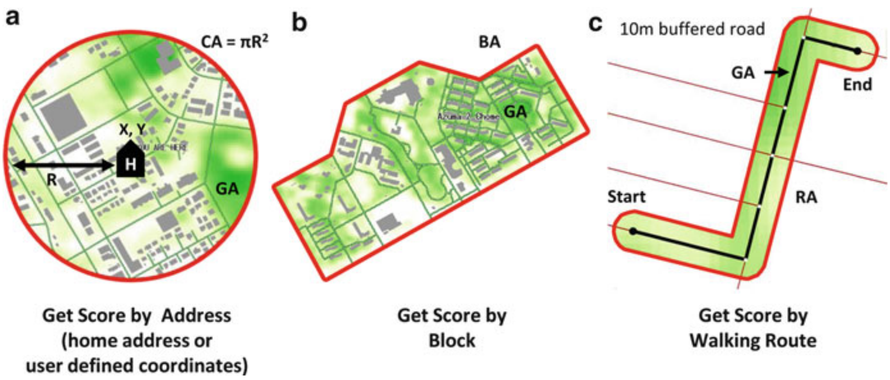


Fig. 16.1 Conceptual view of different greenness score measurements

areas from development. Randall et al. (2003) presented a GIS-based decision support tool to model planning scenarios related to the creation of new green space areas as part of neighborhood greening strategies. Zhang and Wang (2006) presented a study that used landscape metrics to quantify the spatial configurations of green spaces, and performed GIS-based network analyzes to assess the accessibility of many proposed enhancements of green spaces. Ghaemi et al. (2009) implemented a web-based platform "interactive park analysis tool," which is part of the "Green Visions Plan for 21st Century Southern California" project (Wolch et al. 2005). The quality of eco-friendly living places can be measured by an indicator of walkability index or score. Although most walk score calculations are based on distances between home and public facilities, an eco-friendly walk score calculation is based on green spaces (i.e., the location of home or a walking route with green spaces). The higher the score, the better the environmental quality (i.e., eco-friendly) for living or taking green exercise. We propose to measure the greenness score of urban locations through the three modalities illustrated in Fig. 16.1. In the first modality, greenness is measured for the spatial neighborhood around a certain address by user-defined distance (get score by address); in the second modality, greenness is measured for each block of the urban region (get score by block), while the third modality computes the greenness score for walking routes (get score by walking route).

The following equations are used to calculate the greenness score for each of the three modalities (Fig. 16.1a–c).

$$\text{(A) Get score by address} = \left(\frac{\text{GA}}{\text{CA}}\right) \times 100 \tag{16.1}$$

$$\text{(B) Get score by block} = \left(\frac{\text{GA}}{\text{BA}}\right) \times 100 \tag{16.2}$$

$$\text{(C) Get score by walking route} = \left(\frac{\text{GA}}{\text{RA}}\right) \times 100 \tag{16.3}$$

where GA = green area, BA = block area, CA = circle area, RA = 10 m buffered route area, and R = circle radius (user-defined walking distance).

The calculation of a get-score-by-walking route is based on (16.3). This measurement is ideal for informing people who want to make outdoor recreation or exercise activities part of their daily or weekend routines. Outdoor recreation or exercise has become an important element of healthy living, and a remedy against the deficiencies of a modern lifestyle that involves separation from nature. People with special needs, such as the elderly or those with disabilities, often gain therapeutic benefits from activities conducted in a natural environment. Mental well-being is also supported through playing, because play helps establish personal and community identity for children and young people (Bell et al. 2007). A study by Sugiyama et al. (2008) shows that perceived neighborhood greenness was more strongly associated with mental health than with physical health. Moreover, Pretty et al. (2007) summarized the effects on 260 participants of 10 green-exercise case studies (including walking, cycling, horse-riding, fishing, canal boating, and conservation activities) in 4 regions of the United Kingdom. It was determined that green exercise (i.e., exercise in a green area) led to significant improvements in self-esteem and in total mood. The effects were not found to be affected by the type, intensity, or duration of the green exercise.

Furthermore, we can also find the available facilities within a user-defined distance or circle radius based on the accessibility concept. The circle is a spatial analysis boundary whose radius is defined by the user. This radius indicates how far the user is willing to walk to reach the facilities. We also assume that the effectiveness of greenness has a circular pattern known as the distance decay effect. Accessibility is a measure of the spatial distribution of activities around a point location, adjusted for the ability and desire of people to overcome spatial separation (Handy and Niermeier 1997). Several studies describe accessibility, review various accessibility measures, provide case studies, and present novel methods (Bhat et al. 2002; Levinson and Krizek 2005; Thill 2009; Thill and Kim 2005; Handy and Niermeier 1997; Pirie 1979). In short, accessibility describes the ease of travel between a source and a target. For example, having retail stores close to where people live and providing connecting streets increases the likelihood that a person will incorporate walking into their daily routines (Frank and Engelke 2005; Moudon et al. 2007). Furthermore, spatial syntax has been proposed as a new computational language to describe the patterns of modern cities (Hillier 1996; Hillier and Hanson 1984). Typical applications of spatial syntax include pedestrian modeling, crime mapping, and way-finding processes in complex built environments (Hillier 1996; Jiang 1999; Peponis et al. 1990). The axial line-based representation of an urban structure is the earliest approach to spatial syntax (Hillier and Hanson 1984). Recent developments in spatial information science have much to offer for the identification of land-use types, street connectivity, and access to services, in order to determine the factors

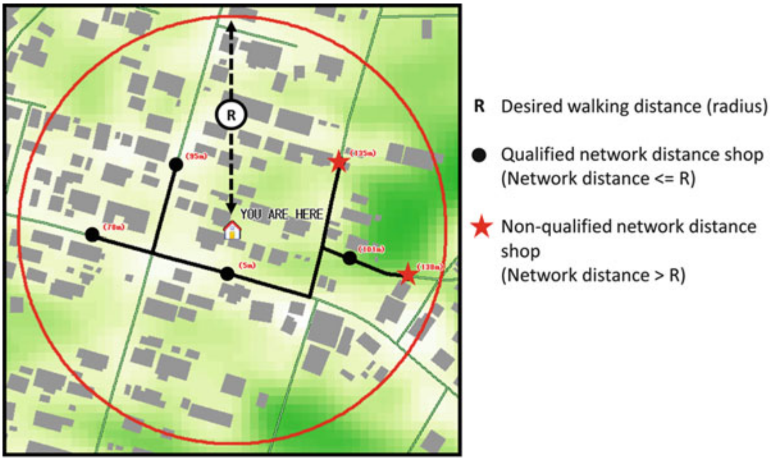


Fig. 16.2 Conceptual view of cumulative opportunity measurements

that might increase or decrease the probability of people being physically active according to selected spatial units of interest (Leslie et al. 2007).

In our urban green space walkability model, users are allowed to choose their desired travel distances and desired types of facility. The desired distance indicates how far people are willing to travel to reach their desired facilities. For example, some may want to reach their desired facilities by walking or cycling, which might mean that they are concerned about the environment, and want to reduce gasoline use and cut carbon dioxide emissions. In contrast, some people are willing to use a car to reach their desired destinations. Moreover, different people may require different facilities. We counted the available facilities based on user choices (i.e., their desired walking distance and desired facility types). The calculation was then based on the "cumulative opportunity," which provides a measure of the number of available facilities within a certain distance or travel time (Fig. 16.2). Examples of cumulative opportunity measures are found in various publications (O'Sullivan et al. 2000; Sherman et al. 1974; Wachs and Kumagi 1973). Cumulative opportunity can be expressed by the following equation:

$$A_i = \sum j(B_j a_j) \tag{16.4}$$

where A_i is the accessibility measured at point i to potential activities in zone j, a_j is the opportunities in zone j, and B_j is a binary or threshold value (i.e., 1 if network distance ≤ search radius, and 0 if network distance > search radius).

We identify all available facilities inside the circle (Fig. 16.2), and then separate them into two groups; one group is composed of a qualified network of facilities for which the travel distance would be less than or equal to the search radius (marked as •), and the other group is the non-qualified network of facilities for

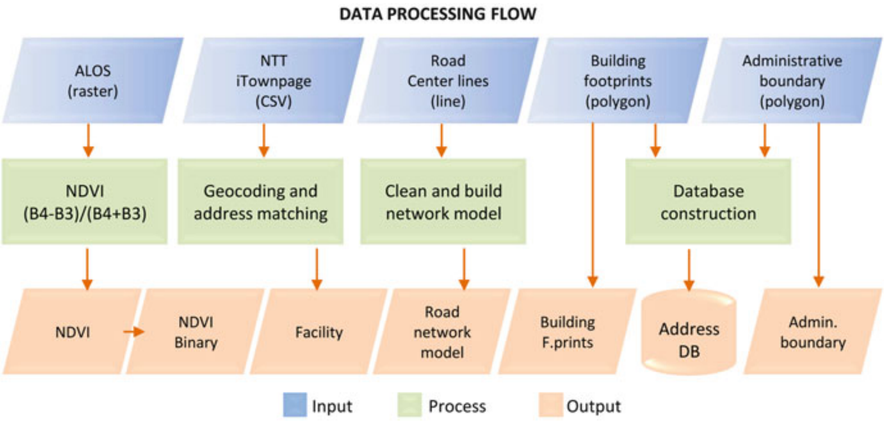


Fig. 16.3 Workflow for data processing and output of data to be used in the eco-friendly walk score calculator

which the travel distance would be greater than the search radius but would still fall within the area of the circle. Moreover, some points, such as shopping malls and supermarkets, include more than one shop. In this model, all measurements between the home and the available facilities are calculated as the actual shortest network distances. Okabe and Okunuki (2001) discussed the advantages of network distances over straight-line distances in the case of a retail market analysis in urban areas. The most traditional analytical tools are based on the assumption that market areas are homogeneous planes, and that the distances can be measured in terms of Euclidean distances. In a small area, however, irregular street layouts produce a heterogeneous plane, and consumers can access stores only through a network of streets. This suggests that there would be great potential demand for analytical tools for micro-spatial analysis of a network in which distance is measured in terms of the shortest route (Fig. 16.3).

16.3 Case Study

16.3.1 Study Area and Data

Our study area was Tsukuba City, a city that was planned for academic and scientific purposes and the home of the University of Tsukuba and the Japan Aerospace Exploration Agency (JAXA). As of 2008, the city had an estimated population of 207,394, and a population density of 730 people per km^2. Its total area is 284.07 km^2. Located approximately 50 km northeast of Tokyo, Tsukuba is sometimes considered part of the Greater Tokyo metropolitan area. Table 16.1 lists the various data used in this study, and their respective uses and sources.

Table 16.1 Data, descriptions, and applications of their use

Data and source	Description	Purpose
ALOS AVNIR-2 (Japan Aerospace Exploration Agency, JAXA)	• Band 3 (red: 0.61–0.69 μm) • Band 4 (infrared: 0.76–0.89 μm) • 10 m spatial resolution at nadir raster in GeoTIFF format	• To delineate green spaces • To convert binary green images • To compute the greenness score
Building footprints (Zmap-TOWNII product from ZENRIN Company)	Building footprints including building name, parcel number, and number of floors Polygon in an ESRI Shape file	• To integrate with administrative boundary data and construct a database of residential addresses • To create masks on vegetated areas
Administrative boundary (Zmap-TOWNII product from ZENRIN Company)	• Administrative boundary including name Polygon in an ESRI Shape file	• To integrate with building footprints and create a database of residential addresses • To calculate the greenness score by administration zone
Road center lines (Geospatial Information Authority of Japan, previously known as the Geographical Survey Institute)	• Road center lines with major road names Line in an ESRI Shape file	• To build a road network model • To measure network distances between a user-defined point and locations of facilities • To compute a greenness score for each road segment • To perform an analysis of the shortest or greenest route
Facility locations (iTownpage from NTT, Nippon Telegraph and Telephone Corp.)	• Business name, address, category, sub-category, business contents, phone number, URL, etc. Comma-separated value (CSV) format	• To convert a point layer for facilities • To find desirable and available facilities by a user-defined search distance

16.3.2 Data Processing

16.3.2.1 Data Processing Workflow

16.3.2.2 Creation of a Binary Green Image from ALOS Satellite Data

ALOS includes an optical sensor known as the advanced visible and near infrared radiometer type 2 (AVNIR-2) with high spatial resolution (10 m at nadir) composed of four multi-spectral bands (i.e., three bands in the visible range and one band in the near infrared region). The normalized difference vegetation index (NDVI; $NDVI = (NIR - RED)/(NIR + RED)$) is computed using a visible red band (RED, Band 3: 0.61–0.69 μm) and a near-infrared band (NIR, Band 4: 0.76–0.89 μm) acquired from vegetation growing seasons. This NDVI (Fig. 16.4) shows the degree of vegetation (intensity) represented as pixel values between 0 and 255, which are stretched from their original values of between −1 and 1.

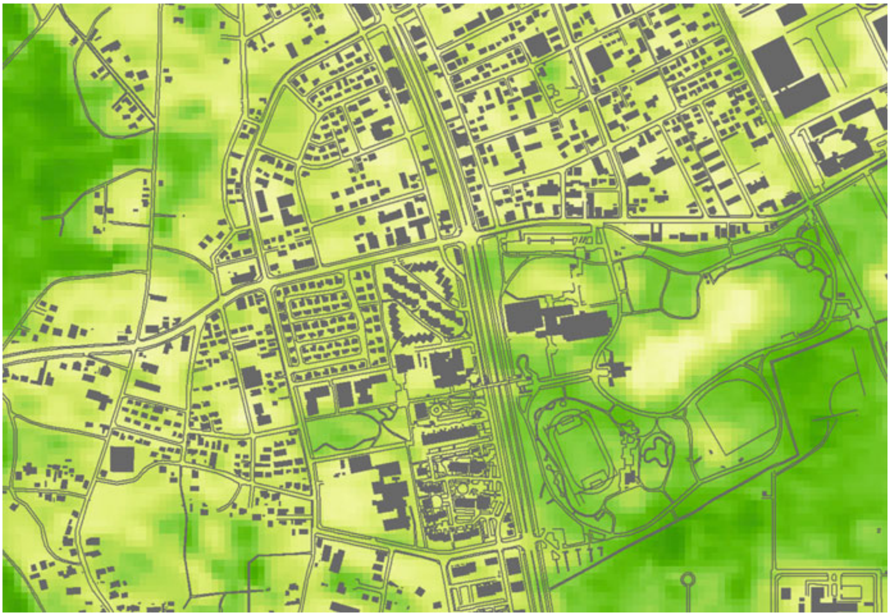


Fig. 16.4 NDVI image (*green intensity*) from an ALOS AVNIR-2 sensor masked with building footprint polygons

After stretching, this NDVI image was re-sampled to a 5 m spatial resolution with the Spline interpolation method using the ArcGIS Spatial Analyst extension (i.e., we converted the 10 m raster to a 10 m point, and the 10 m point to a 5 m raster again, by setting the output raster resolution at 5 m). The purpose of this re-sampling was to reduce the errors between raster and vector analysis (Fig. 16.5) because the analysis is made between the 10 m road-buffer line and the center point of each raster. Another way to convert low resolution to high resolution is called "pan-sharpening." However, this process requires an additional high-resolution panchromatic band. The use of pan-sharpened images in the analysis of vegetation is very rare because the original spectral values are transformed during the pan-sharpening process. Pan-sharpened images are commonly used for visualization purposes. We use original spectral properties to calculate the NDVI and interpolate the middle one. Interpolation of the image is a technique in which the spatial resolution of an image is increased from its original size to a higher resolution to improve the image quality.

To separate vegetated and non-vegetated spaces, we set the threshold at 113 of the NDVI pixel values by comparing two images (i.e., one from the 67 cm RGB-321 true color ortho-image and one from the 5 m re-sampled ALOS NDVI images) using the view > link/unlink viewers > geographical function in the commercial remote sensing software ERDAS Imagine (Fig. 16.6). After this step, the intensity image is converted into a binary green image (1 for vegetated

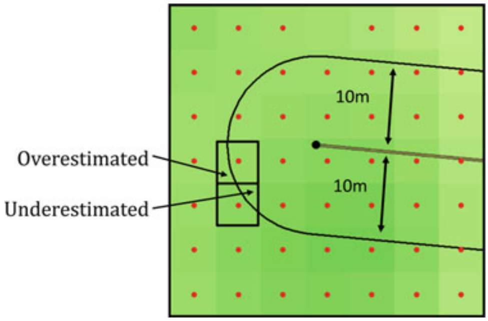


Fig. 16.5 Errors between raster cells and vector line analysis, depending on cell size

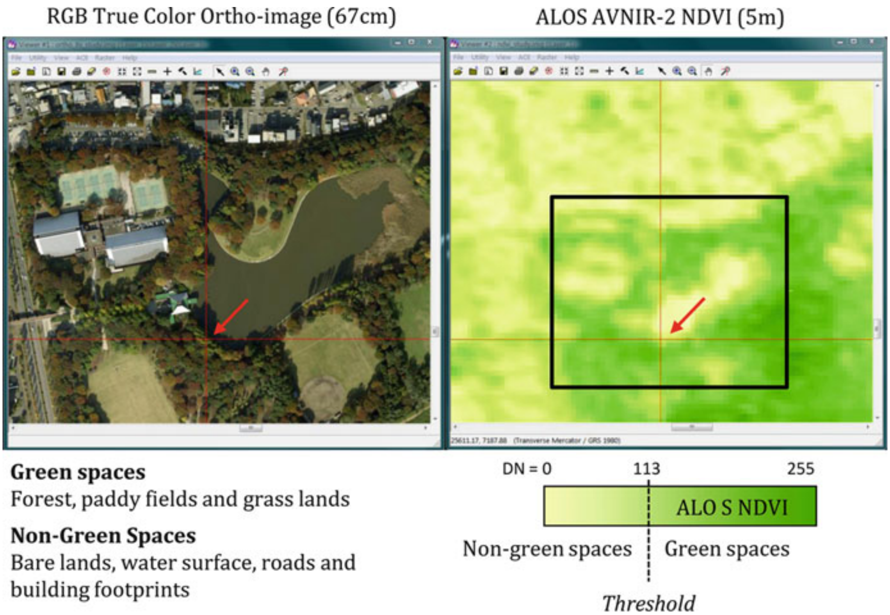


Fig. 16.6 Identification of the threshold value by linking the 67-cm ortho-image and the 5 m re-sampled ALOS NDVI image (viewing the actual landscape features from the high-resolution ortho-image *left view*, and obtaining the NDVI pixel values from the *right view* while moving the cross-hairs by using ERDAS commercial remote-sensing software)

area, and 0 for non-vegetated area). The main purpose of this conversion is to identify the vegetated areas rather than the vegetation intensities, which vary from season to season. The binary green image also reduces the data size and the required computational time. This procedure is especially suitable for web-based GIS in which the network and computational resources are limited. Vegetated areas included trees, bush land, grass land, and paddy fields. Non-vegetated areas include buildings, parking lots, bare land, rivers, and lakes.

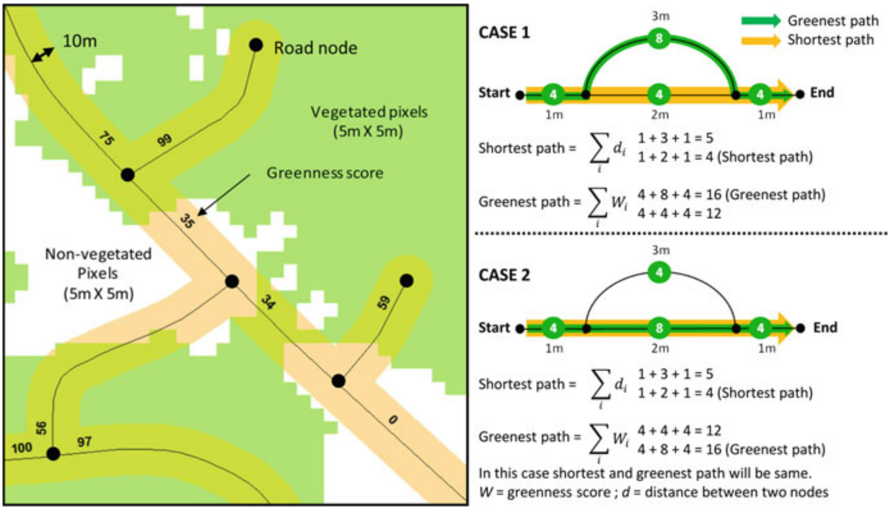


Fig. 16.7 Calculation of the greenness score for each road segment based on the binary green image, and an example of the shortest and greenest path calculations using greenness score as the weighted factor

16.3.2.3 Computation of the Greenness Score and the Road Network Model

Road center-line data were acquired from the Geospatial Information Authority of Japan. However, this data set does not cover all the small streets in the city. Therefore, small streets that were missing were digitized based on Zmap-TOWNII data. Following this, we added a 10 m buffer to both sides of the road, and computed the greenness score based on the binary green image (Fig. 16.7) for each road segment using (16.3). Next, we built a topological road network model using VDS road network builder provided by VDS technologies. In this process, we set up the greenness score attribute field as a weight factor in order to compute the shortest or greenest route between two points. The shortest route was computed based on road distance, while the greenest route was computed based on road distance and the greenness score, whose value ranges from 0 to 100.

16.3.2.4 Construction of a Residential Address Database

For this case study, a database of residential addresses was created from a combination of administrative boundary and building footprint data sets. Building footprint data sets are useful for estimating building populations (Lwin and Murayama 2009) because such data contain rich attributes including building number, number of floors, and building name. Unlike other countries, most Japanese addresses are based on a block-by-block system. The address does not contain a street or road name; instead, it is expressed by a sequence of blocks (prefecture block, city block, ward block, ownership block, etc.). For this study, we constructed the address database by performing an intersection function between these block layers. We separated the addresses into two parts: the main block and the sub-block. The main block represented the smallest administrative unit, and the sub-block represented the smallest land unit. For example, in the case of Kasuga 3–15-23, Kasuga 3 was constructed from an administrative boundary block, and 15-23 was constructed from the smallest land unit. The purpose of the address database is to locate the place in a user-friendly way, and to avoid problems with mis-typing when performing an address search. Although this approach is not appropriate for large land blocks (e.g., factories, schools, and hotels), users can still locate their position and the distance of the desired destination by using the interactive map circle tool.

16.3.2.5 Conversion of Public Facility Data

Our model also uses the count of available facilities classified by a user-defined search area and specified facility types. This is useful for potential home-buyers and current residents to calculate the distances between home and available facilities on the network. We use iTownpage data, which were downloaded from the Nippon Telegraph and Telephone Corp. (NTT) website. These data include the business name, type, category, content, address, telephone number, and other information in a comma-separated value (CSV) format. The iTownpage website supports the everyday life and business activities of visitors and expatriates in Japan, as well as people living overseas, by enabling users to search for information about stores and businesses via the Internet. These CSV data were converted into ESRI point features (Fig. 16.8) using commercial geo-coding software with an accuracy at the building level. These NTT iTownpage data can be used to separate the residential and non-residential buildings, and to carry out other retail market analysis. For example, Lwin and Murayama (2010) used NTT iTownpage data to separate the building-use types and integrate those with a digital volume model (DVM) that was derived from light detection and ranging (LiDAR) data to produce a fine-scale dasymetric map for Tsukuba City.

16.3.3 Implementation of the Eco-Friendly Walk Score Calculator

We have implemented a system called the “eco-friendly walk score calculator” based on our urban green space walkability model. The overall system is built on Microsoft ASP.NET with an AJAX extension and VDS technologies (web mapping components for ASP.NET). ASP.NET is a web application framework marketed by Microsoft that programmers can use to build dynamic websites, web applications, and XML web services. AJAX (shorthand for asynchronous JavaScript and XML)

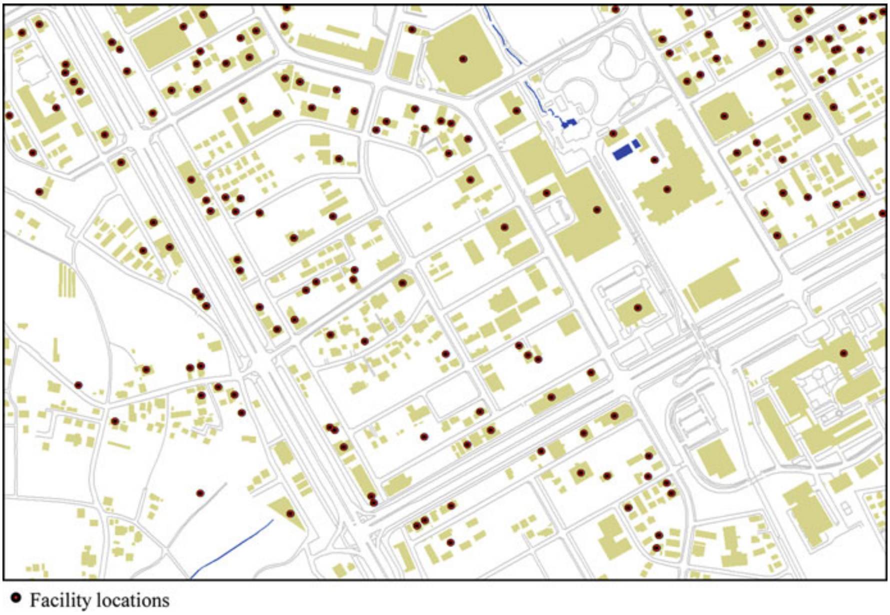


Fig. 16.8 Conversion of an iTownpage comma-separated value (CSV) file to ESRI point features (one point may contain many shops, such as a shopping mall and supermarkets)

is a group of interrelated web development techniques used on the client side to create interactive web applications. With AJAX, web applications can retrieve data from the server asynchronously in the background without interfering with the display and behavior of the existing page. The use of AJAX techniques has led to an increase in interactive and dynamic interfaces on web pages. AspMap for .NET from VDS technologies is a set of high-performance web-mapping components and controls for embedding maps in ASP.NET applications (web forms). Figure 16.9 shows the overall system design and potential users.

16.4 Model Outcomes

Figure 16.10 shows the graphical user interface (GUI) of the eco-friendly walk score calculator. We measured the greenness score via the three modalities called "get score by address," "get score by block," and "interactive score." The first measurement mode of our program is the "get score by address" function (Fig. 16.11). This is ideal for current residents to evaluate the environmental quality of their neighborhood by entering their home address and search radius (the default search radius is 250 m). This tool also finds the available facilities within the user-defined search radius based on the concept of accessibility. The second measurement mode of our program is the "get score by block" (Fig. 16.12) function, which is ideal for

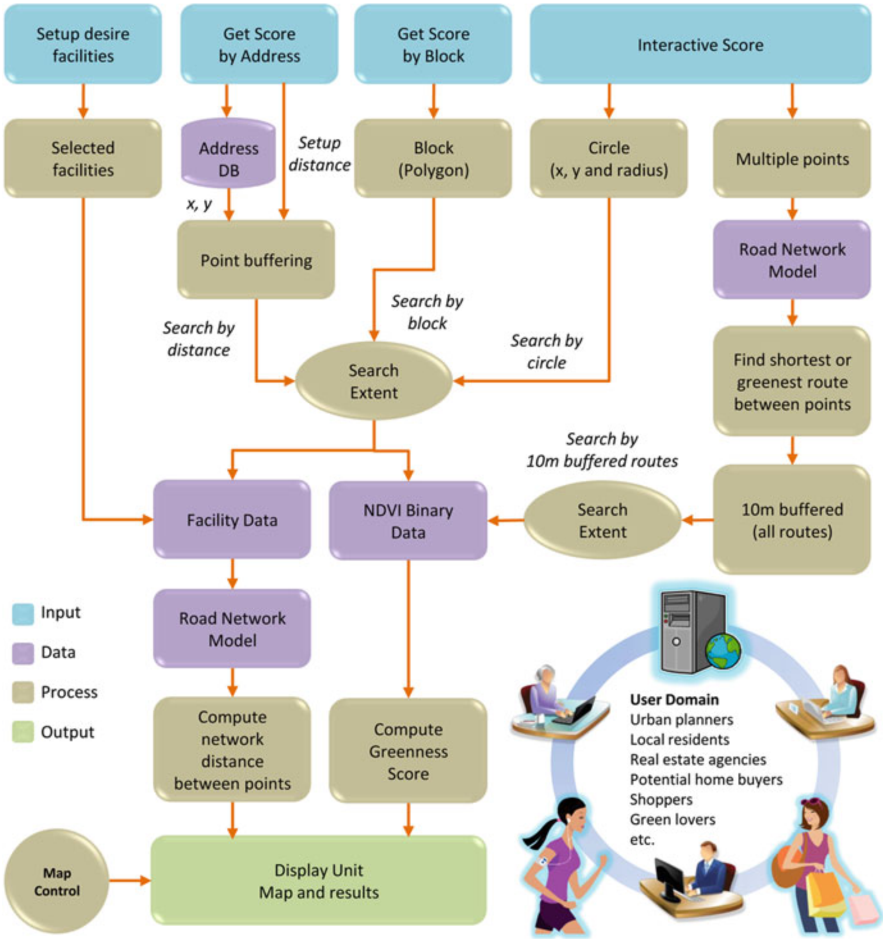


Fig. 16.9 System design of the eco-friendly walk score calculator

urban planners to evaluate the greenness of spaces according to planning zone (i.e., administrative block). The third measurement mode of "interactive score" is ideal for potential home-buyers who are planning to live in Tsukuba City, or for local residents who want to walk along the greenest or shortest route between locations. As for potential home-buyers, users can locate their location and desired walking distance by drawing a circle on a map (Fig. 16.13).

The calculation of the greenness score is the same as for the get score by address mode. In interactive score mode, users can also find either the shortest or the greenest walking route (Fig. 16.14) by specifying their start and end points. The shortest route is ideal for shopping activities, and the greenest route is ideal for recreational

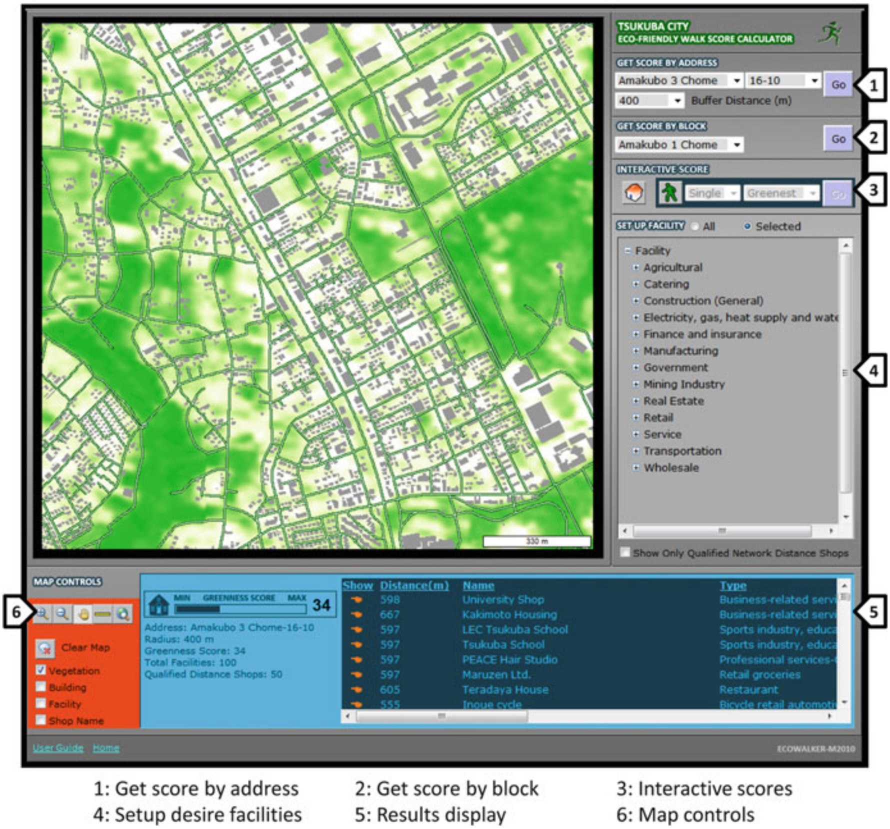


Fig. 16.10 Graphical user interface of the eco-friendly walk score calculator (URL: http://land.geo.tsukuba.ac.jp/ecowalkscore)

walking activities. Finally, users can also compute the walking score for multi-stop trips. To this end, the user specifies multiple activity sites. For example, one could start from home, go to the library, continue on to a shopping center, and then return home (Fig. 16.15).

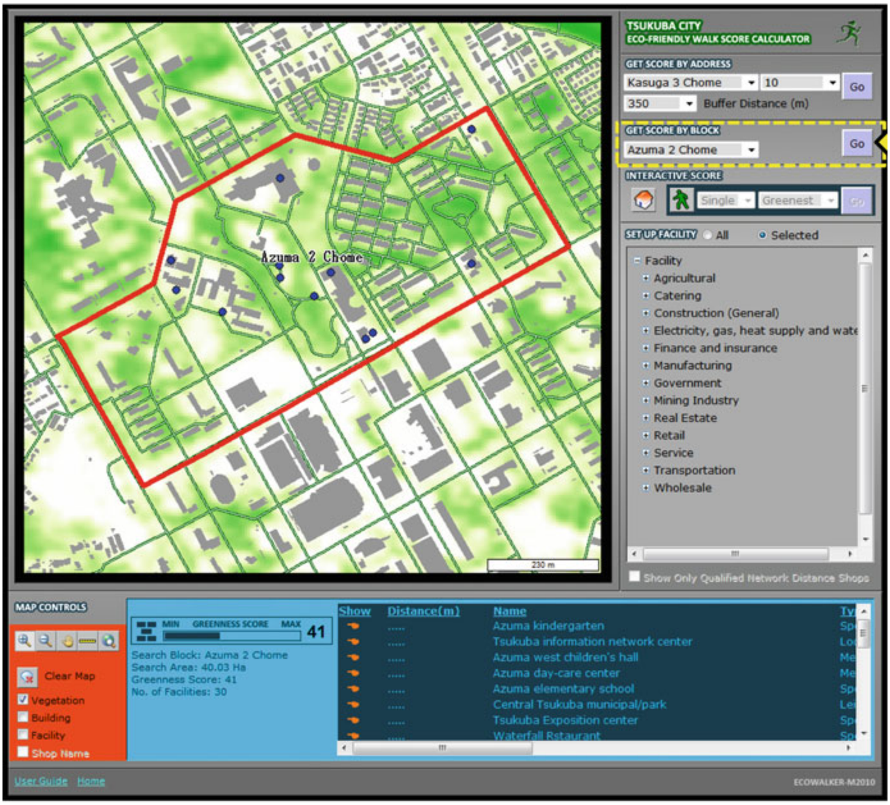
16.5 Qualitative Usability Study

In order to evaluate our web-based GIS eco-friendly walk score calculator in Tsukuba City, we conducted a number of face-to-face interviews and telephone conversations with university students, researchers, private companies, non-profit organizations (NPOs), local residents, real-estate agencies, and city planners. Of the groups of users that were part of this study, real-estate agents found the ability to

Fig. 16.11 Get score by user-defined address and default search radius of 250 m (*Note*: one point may contain more than one shop)

show the neighborhood environmental quality and surrounding public facilities to potential home buyers a highly valuable resource, while local residents tended to favor the get score by address function. Students, on the other hand expressed a preference towards the get score by route analysis capabilities.

In order to provide an avenue for system improvement, researchers proposed incorporating the ability to integrate additional scores or indices to the system, including accessibility and connectivity indicators such as the alpha or beta index, average block length, or average block size based on a GIS network data model (Thill 2000). City planners in our usability study suggested that in addition to the get score by administrative block function, scores could also be calculated by irregular boundaries such as a user-defined polygon, given that land-use planning is often performed according to land-use type or within specific properties. For example, computing the greenness score for a university would help to improve campus environmental quality.



Search Block: Azuma 2 Chome ; Search Area: 40.03 Ha ; Greenness Score: 41 ;
No. of Facilities: 30

Fig. 16.12 Getting a score by administrative block

NPOs found the system to be moderately useful, since their primary purpose is to locate open spaces for humanitarian assistance and other social or cultural activities. Although we did not find any significant difficulties with the graphical user interface, a handful of students commented on the size of the GUI panel, suggesting that the eco-friendly walk score calculator would be handy to have as an application on their smart phones or Netbook computers in order to find the greenest route while they walk or exercise. They also suggested making separate route analysis web-GIS pages for mobile Internet users. Overall, the system was evaluated favorably by real-estate agencies, researchers, and students.

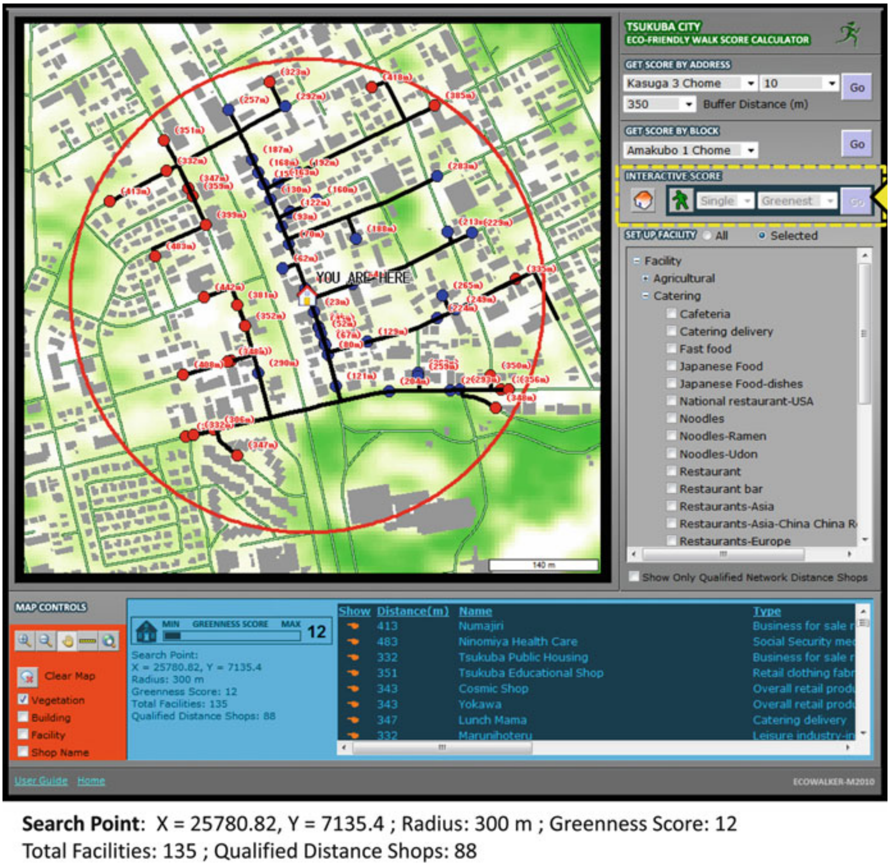


Fig. 16.13 Interactive score using the circle tool for potential home-buyers

16.6 Conclusion

The increasing popularity of the Internet and user-friendly web-based GIS applications such as Google Maps/Earth and the Microsoft Bing Maps platform have made GIS an integral part of life today for finding the nearest facilities, driving routes, and so on. However, choosing an eco-friendly place to live or a walking route is a big challenge for local residents because of the lack of GIS analytical functions and environmental data available online. Although the analysis of route paths has been widely used in GIS applications, the integration of green factors with the analysis of the route path is still lacking in the GIS arena. In this chapter, we have presented an integrated methodology for identifying an eco-friendly place to live or to walk by providing web-based GIS analytical functions using Tsukuba City in Japan as a case

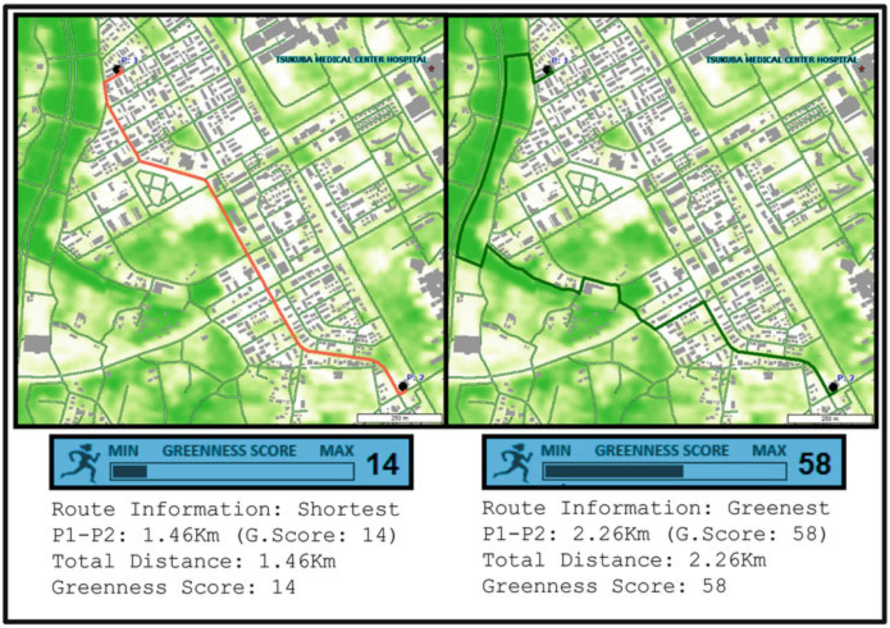


Fig. 16.14 Finding walking routes by either the shortest route or the greenest route

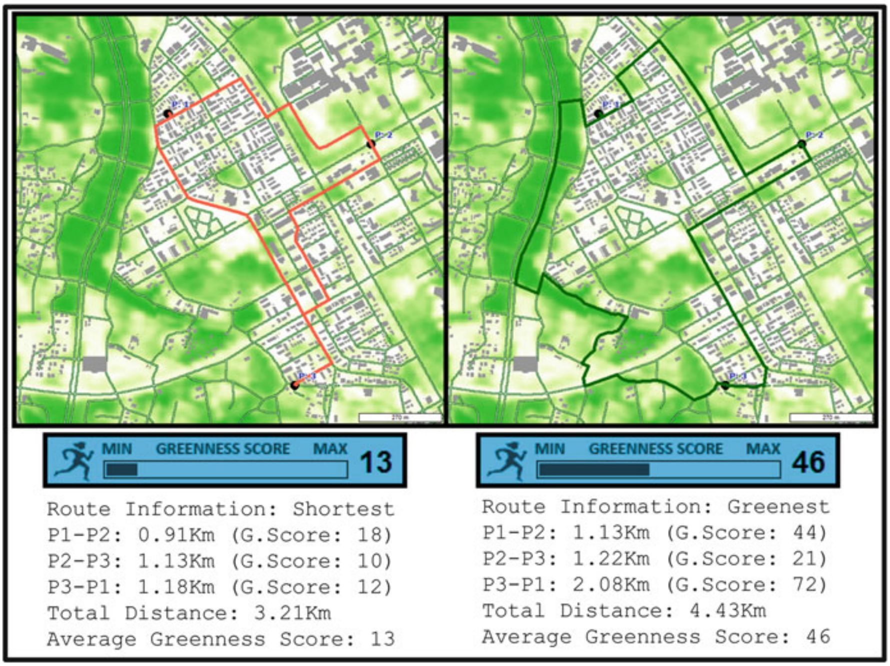


Fig. 16.15 Finding multiple places (points) with a walking route through use of the greenness score

study. This web-based, eco-friendly walk score calculator enables users to evaluate the environmental quality of a neighborhood, to find the nearest facilities which are accessible on foot, to choose an eco-friendly place to live for potential home-buyers, and to choose a route for green exercise. Although this web-based GIS represents a fairly localized prototype, we hope it will help local city planners and policy-makers to build sustainable eco-cities to improve the mental and physical health of their residents in various parts of the world.

References

Abley S (2005) Walkability scoping paper. Charted Traffic and Transportation Engineering, Christchurch

Bell S, Tyrvainen L, Sievanen T, Proebstl U, Simpson M (2007) Outdoor recreation and nature tourism: a European perspective. Living Rev Landsc Res 1:2

Bhat C, Handy S, Kockelman K, Mahmassani H, Gopal A, Srour I, Weston L (2002) Development of an urban accessibility index: formulations, aggregation, and application. Report No. FHWA/TX-02-4938-4. Center for Transportation Research, The University of Texas at Austin, Austin

Dizdaroglu D, Yigitcanlar T, Dawes LA (2009) Sustainable urban futures: an ecological approach to sustainable urban development. In: Proceedings of the second infrastructure theme postgraduate conference 2009: rethinking sustainable development planning, infrastructure engineering, design and managing urban infrastructure, Queensland University of Technology, Brisbane

Forsyth A, Southworth M (2008) Cities afoot: pedestrians, walkability, and urban design. J Urban Des 13:1–3

Frank LD, Engelke P (2005) Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution. Int Reg Sci Rev 28:193–216

Ghaemi P, Swift J, Sister C, Wilson JP, Wolch J (2009) Design and implementation of a web-based platform to support interactive environmental planning. Comput Environ Urban Syst 33:482–491

Grahn P, Stigsdotter UA (2003) Landscape planning and stress. Urban For Urban Green 2:1–18

Handy S, Niermeier DA (1997) Measuring accessibility: an exploration of issues and alternatives. Environ Plann A 29:1175–1194

Hillier B (1996) Space is the machine: a configurational theory of architecture. Cambridge University Press, Cambridge

Hillier B, Hanson J (1984) The social logic of space. Cambridge University Press, Cambridge

Jiang B (1999) SimPed: simulating pedestrian crowds in a virtual environment. J Geogr Inform Decis Anal 3:21–30

Kweon B-S, Sullivan WC, Wiley AR (1998) Green common spaces and the social integration of inner city older adults. Environ Behav 30:832–858

Leslie E, Coffee N, Frank L, Owen N, Bauman A, Hugo G (2007) Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health Place 13:111–122

Levinson DM, Krizek KJ (2005) Access to destinations. Elsevier, Kidlington

Lwin KK, Murayama Y (2009) A GIS approach to estimation of building population for micro-spatial analysis. Trans GIS 13:401–414

Lwin KK, Murayama Y (2010) Development of a GIS tool for dasymetric mapping. Int J Geoinform 6:11–18

Maas J, Verheij RA, Groenewegen PP, de Vries S, Spreeuwenberg P (2006) Green space, urbanity, and health: how strong is the relation? J Epidemiol Comm Health 60:587–592

Mahon JR, Miller RW (2003) Identifying high-value green space prior to land development. J Arboricult 29:25–33

Moudon AV, Lee C, Cheadle AD, Garvin CW, Johnson DB, Schmid TL et al (2006) Operational definitions of a walkable neighborhood: theoretical and empirical insights. J Phys Act Health 3:S99–S117
Moudon AV, Lee C, Cheadle AD, Garvin C, Johnson DB, Schmid TL et al (2007) Attributes of environments supporting walking. Am J Health Promot 21:448–459
O'Sullivan D, Morrison A, Shearer J (2000) Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. Int J Geogr Inform Sci 14:85–104
Okabe A, Okunuki K (2001) A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. Trans GIS 5:209–220
Owen N, Humpel N, Leslie E, Bauman A, Sallis JF (2004) Understanding environmental influences on walking: review and research agenda. Am J Prev Med 27:67–76
Peponis J, Zimring C, Choi YK (1990) Finding the building in way-finding. Environ Behav 22:555–590
Pirie GH (1979) Measuring accessibility: a review and proposal. Environ Plann A 11:299–312
Pretty J, Peacock J, Hine R, Sellens M, South N, Griffin M (2007) Green exercise in the UK countryside: effects on health and psychological well-being, and implications for policy and planning. J Environ Plann Manag 50:211–231
Randall TA, Churchill CJ, Baetz BW (2003) A GIS-based decision support system for neighbourhood greening. Environ Plann B 30:541–563
Reid S (2008) Fit for purpose: evaluating walkability. Eng Sustain 161:105–112
Roehr D, Laurenz J (2008) Green surfaces in the city context. In: Proceeding of ecocity world summit, San Francisco
Sallis JF, Frank LD, Saelens BE, Kraft MK (2004) Active transportation and physical activity: opportunities for collaboration on transportation and public health research. Transport Res Pol Pract 38:249–268
Sherman L, Barbara B, Kondo W (1974) Method for evaluating metropolitan accessibility. Transport Res Rec 499:70–82
Sugiyama T, Leslie E, Giles-Corti B, Owen N (2008) Associations of neighbourhood greenness with physical and mental health: do walking, social coherence and local social interaction explain the relationships? J Epidemiol Comm Health 62:e9
Thill J-C (2000) Geographic information systems for transportation in perspective. Transport Res C Emerg Tech 8:3–12
Thill J-C (2009) Transportation applications of geographic information systems, manual of geographic information systems. In: Madden M (ed) Manual of geographic information systems. ASPRS, Washington, pp 1035–1049
Thill J-C, Kim M (2005) Trip making, induced travel demand, and accessibility. J Geogr Syst 7:229–248
Wachs M, Kumagi TG (1973) Physical accessibility as a social indicator. Soc Econ Plann Sci 7:437–456
Wolch J, Wilson JP, Fehrenbach J (2005) Parks and park funding in Los Angeles: an equity-mapping analysis. Urban Geogr 26:4–35
Zhang L, Wang H (2006) Planning an ecological network of Xiamen Island (China) using landscape metrics and network analysis. Landsc Urban Plann 78:449–456

Chapter 17
Watershed Evaluation Using Geospatial Techniques

Kondwani Godwin Munthali and Yuji Murayama

17.1 Introduction

Soil erosion and the sedimentation of reservoirs are serious problems throughout the tropics, and result from severe and uncontrolled environmental degradation. As a result, declining watershed resources continue to put great pressure on the available agricultural land to support households as soil erosion increases, leading to considerable loss of soil fertility and in extreme cases to eventual desertification (Munthali et al. 2011). In the developing regions, cut-and-burn agricultural practices have been identified as the main driver of erosion, and they pose a great risk to the ecosystems to which such watersheds belong (Chimphamba et al. 2006). Physiologically, many tropical river regimes are unstable and pose a great danger to life and infrastructure as they continuously meander and change course (Munthali et al. 2011). Seasonally, it is estimated that tropical floods inundate significant proportions of fertile land (Norplan A.S. in Association with COWI et al. 2003; WWF (World Wide Fund for Nature) 2009).

It has been established that of all the water-related problems in the tropics, the sedimentation of reservoirs is one of the most economically crippling because of the large investments in dams for hydroelectric power and irrigation (Nagle et al. 1999). With the world's dams costing billions, replacing lost reservoir storage is a costly endeavor even with reported rough estimates of lost capacity being as little as 1% annually (Mahmood 1987). Despite a number of programs being proposed and implemented to control erosion and slow sedimentation, observers have argued that many of these efforts are doomed to failure (Magrath and Doolette 1990; Mahmood 1987). Therefore, the important questions to ask are whether soil erosion

K.G. Munthali (✉) • Y. Murayama
Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
e-mail: kmunthali@gmail.com; mura1@sakura.cc.tsukuba.ac.jp

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_17,

reduction, and hence sediment reduction, can be planned and executed effectively, and secondly when and where the watershed can actually be managed (Nagle et al. 1999). This is not only because sediment movement in watersheds is a complex process that takes place over long periods of time, but also because of the understanding that while soil conservation programs that focus on agricultural lands are necessary and useful for many other reasons, these may not be the key problem areas as far as the control of sedimentation is concerned (Nagle et al. 1999). This highlights the key weakness of most sediment management programs in the tropics, i.e., their limited capacity to identify and focus efforts on key problem areas (Nagle et al. 1999). This can be blamed in part on the lack of reliable data on soils and erosion rates on the one hand, and limited financial and/or human resources on the other (Munthali et al. 2011).

However, with advances in remote sensing technology and the development of geographical information systems (GIS), watershed evaluation has been given a significant boost. Nagle and others (Nagle et al. 1999) contend that coupled with good field studies that identify sediment sources, the sophistication of geospatial techniques should provide significant improvements in watershed evaluation and management. This is even more important when pressing economic needs for irrigation water and hydroelectric power in the developing tropical regions force many dam projects to be implemented with the scantiest consideration to the potential for rapid sedimentation (Dunne 1988). Because watershed management involves flood and sedimentation control as well as soil erosion management in tropical regions, the rest of this chapter begins by providing a background to watershed evaluation, focusing on methodological theories of geospatial watershed evaluation relating to soil erosion modeling. Without loss of generality, we then present a geospatial watershed evaluation case study of the Songwe River watershed in Northern Malawi, and finally discuss the usefulness of the techniques considered.

17.2 Background to Watershed Evaluation

17.2.1 Flood Control and Sedimentation

There are significant differences in the precision of regulated and natural flows between individual dams, as well as between their methods of operation (Munthali et al. 2011). However, common to them all is the fact that the construction of a dam does affect the river's distribution of discharge and its suspended material (Wellmeyer et al. 2005). It is known that meandering river-flow regimes follow a common evolutionary pattern over time (Friedkin 1945). Therefore, the key to regularizing the flow is to alter the natural discharge regime that subsequently affects the behavior, or the rate of meander and erosion evolution, and hence sedimentation (Wellmeyer et al. 2005).

Environmentally, sediment transportation reflects the distributed erosion processes acting in the basin, as well as being a measure of the slow process of degradation and sequential loss of one of agriculture's critical natural resources, top soil (Pilotti and

Bacchi 1997). Erosion processes are generally slow (Munthali et al. 2011), but the annual volumes of the sediment load involved are huge (U.S. Department of Agriculture-Soil Conservation Service 1994). Sediment load is defined as debris eroded from an area drained by a stream, which is delivered to and transported by the stream itself (Schumm 2009). The total amount of erosional debris exported from such a drainage basin is its sediment yield (Munthali et al. 2011).

The sediment load is usually the rock underlying the drainage basin. Until this rock is weathered into transportable fragments the sediment yield is low, which prompted Schumm (2009) to point out that an evaluation of the erosion conditions provides a good estimate of the sediment yield. It has been further argued that a more reliable assessment of watershed sediment yield is obtained when a hydraulic approach is augmented by information about the mechanisms that feed sediments into the channeled flow (Pilotti and Bacchi 1997).

Disturbances in vegetative cover (land use/cover changes due to agriculture, timber, and charcoal harvests, construction, and others) leave soils vulnerable to erosion (Munthali et al. 2011). However, as Fried et al. (2000) observed, where this happens and how much is eroded and transported downstream largely depends on the topography and the hydrological properties of the soil. Once eroded, the sediment's journey down the catchment area depends significantly on runoff, which reduces in speed over time and then results in deposition (Fried et al. 2000). Because of the large amount of human activity impacting on river flow, many dams have been constructed across river systems for various purposes, which include flood control, recreation, and power generation, and these have played a major role in determining the sites of sediment deposition (Munthali et al. 2011). Many of the reservoirs end up holding much of the sediment load themselves (Haregeweyn et al. 2006; Tamene et al. 2006; Schumm 2009).

Bozali et al. (2008) and Pandey et al. (2008) highlight the fact that the success of reservoir construction works to engineer stable river courses and control floods rests in the sustainability of the watershed itself. As noted, the key weakness of most sediment management programs in the tropics has been their limited capacity to identify and focus efforts on key problem areas (Nagle et al. 1999). They (Nagle et al. 1999) continue to advise that if the control of reservoir sedimentation is the principal reason for such programs, more critical thought must be given to the description and quantification of major sediment sources. In other words, a geospatial watershed evaluation should, among other issues, try (1) to quantify the hydrological sediment potential, and (2) to determine critical areas requiring prioritized conservation management in that particular watershed (Munthali et al. 2011).

17.2.2 Soil Erosion Management

Many of the answers to whether a watershed is sustainable or not depend on how much of the soil erosion has been contained. Erosion by running water resulting from precipitation is the most severe hazard threatening the protection of watersheds from soil loss (Pandey et al. 2008; Commission of the European Communities 2006). Over time, such erosion-induced changes to the drainage basin affect river discharges (Mulder and Syvitski 1996). It is therefore widely understood and accepted that an evaluation of the state of erosion needs an objective methodology, operating on standard data sets, which allows the assessment to be repeated as conditions, pressures, and drivers change, or the broad-scale implications of prospective watershed changes to be explored (Kirkby et al. 2008). In the early days of erosion modeling, the focus was on a broad-scale approach that could readily be applied in a wide range of conditions to give advice on conservation practices (see USLE, Wischmeier and Smith 1958, 1978; and RUSLE, Renard et al. 1991). However, much recent work (for example, WEPP, Nearing et al. 1989; EUROSEM, Morgan et al. 1994; KINEROS, Smith et al. 1995; LISEM, de Roo 1996; PESERA, Kirkby et al. 2008) has concentrated on detailed process-based modeling (Kirkby et al. 2008).

The USLE model developed from the desire to keep the erosion of cultivated fields within acceptable limits given the climate, slope, and agricultural production factors of a particular region (Roose 1996). It is a multiplier model in which if one factor tends toward zero the estimated erosion will also tend toward zero (Roose 1996), with *A*, the potential long-term average annual soil loss in tons per acre per year, given as

$$A = R \times K \times \mathrm{SL} \times C \times P \tag{17.1}$$

where R is the rainfall erosivity index based on the rainfall and runoff factor of that particular region, K is the soil erodibility factor that depends on the organic matter and texture of the soil, its permeability, and profile structure, SL is a topographic factor derived from the regional length and gradient of slopes, C is the plant cover, which is a simple relation between erosion on bare soil and erosion observed under a cropping system, and P is any specific erosion control practices factor such as contour ridging (Roose 1996).

USLE has been a practical model for finding rational solutions to practical problems, especially in areas where data availability is a challenge. However, because of its intrinsic limitations (see Roose 1996), variant forms of it have appeared, each trying to fix the limitations and adapt the model to particular geographic locations (see RUSLE, Renard et al. 1991; SLEMSA, Elwell 1981; and EPIC, Williams 1982). In addition, uncertainty still hangs over the universal applicability of the USLE and its variant forms in many other areas, especially regions where landslides and linear erosion are predominant (Roose 1996). In particular, Kirkby et al. (2008) notes that the USLE-approach fails to distinguish properly between soil and climatic conditions in the infiltration process.

Process models, on the other hand, have been reported to have the potential to respond explicitly and in accordance with experience to changes in climate or land use, and so show great promise for developing scenarios of change and "what-if" analyzes of policy or economic watershed options (Kirkby et al. 2008). It is argued that a process-based approach, as opposed to the USLE approach, does (1) apply the same objective criteria to all areas, and so can be applied throughout a region, subject to the availability of suitable generic data, (2) provide a quantitative estimate of erosion rate, which can be compared with long-term averages for tolerable erosion, and (3) allow the methodology to be re-applied with equal consistency as available data sources are improved, and for past and present scenarios of changing climate and land use (Kirkby et al. 2008). It should be mentioned, though, that there are a number of possible methodological approaches for creating erosion assessment maps for watersheds (Gobin et al. 2004), each of which has its own strengths and weakness, and therefore a choice of one over the other depends mainly on the objective and the prevailing conditions of the particular watershed.

17.3 Songwe River Watershed Geospatial Evaluation: A Case Study

17.3.1 Case Study Area

The case study area is a 4278 km^2 watershed, with the 200-km Songwe River forming a physical boundary between the United Republic of Tanzania and the Republic of Malawi. It is part of the semi-arid eastern and southern African Great Rift Valley, and lies between latitudes 9°6′23″–9°56′17″ south and longitudes 32°44′34″–33°56′31″ east. Slightly over half of its total area falls in the districts of Ileje, Mbozi, Mbeya, and Kyela of the Mbeya region on the Tanzanian side, and the rest is in Karonga and Chitipa districts on the Malawi side (Munthali et al. 2011).

17.3.2 Methods and Database

The pan-European soil erosion risk assessment (PESERA), a simplified hydrological process-based approach, was used in this study to combine hydrological surface runoff factors in order to estimate sediment yield (Munthali et al. 2011). By attempting to use advances in the understanding of runoff processes as opposed to sediment transport, it was only sensible for a forecast of runoff and soil erosion in PESERA to be built on a hydrological core (Kirkby et al. 2008). While this study concentrated on calibration, validation, and its scenario application in the tropical region of the study area (Munthali et al. 2011), a development account of the PESERA model is provided by Kirkby et al. (2008).

With sedimentation being sensitive to both climate and land use, as well as to detailed conservation practices, temperature and precipitation proxies were used to estimate the climatic parameters (Munthali et al. 2011). Evapo-transpiration can be obtained empirically, but in the case of PESERA, it is partitioned proportionally to vegetative crown cover (Kirkby et al. 2008). Its potential was determined using the *Hargreaves model* (Munthali et al. 2011).

The monthly rainfall was derived by averaging the monthly daily rainfall data for the years 1998 through 2006, which was obtained from seven gauging stations from which monthly mean rain per rainy day and mean monthly rains were derived for each gauging station and interpolated to fit the entire watershed (Munthali et al. 2011). The infiltration excess overland flow runoff was estimated from storm rainfall and soil moisture. A Hortonian process of point hydrological balance was then used to estimate sediment transport from the excess overland flow (Munthali et al. 2011). This was coupled to a vegetation growth and soil model in order to (1) budget for the living biomass and organic matter subject to the constraints of land use and cultivation choices, and (2) estimate the required hydrological variables from moisture, vegetation, and seasonal rainfall history, respectively (Kirkby et al. 2008).

Runoff thresholds determine near-surface water storage, and soil properties constrain its upper limit (King et al. 1995). The model parameters of available water storage capacity, crustability, and erodibility were therefore obtained from the available soil properties (Munthali et al. 2011). For agricultural land, full water storage after ploughing decays exponentially with time, and reduces to a minimum in vegetated areas (Darboux et al. 2002; Le Bissonnais et al. 2005). This allows for a seasonal response in runoff thresholds, and therefore in infiltration excess overland flow (Kirkby et al. 2008).

Vegetation cover reduces with real-time processes of fire, plant gathering, and grazing (Kirkby et al. 2008). This relates empirically to the seasonal cover cycle and/or above-ground biomass of the land-use classes (Haboudane et al. 2002). Therefore, the PESERA model estimated sub-surface flow, which influences the infiltration excess overland flow, using the *TopModel* (Beven and Kirkby 1979). Topographic properties were estimated from local relief maps (Munthali et al. 2011).

Sediment transport was then estimated by sediment yield, Y (kg m^{-2} year^{-1}), defined as the sediment transported to the slope base, averaged over the slope length (Kirkby et al. 2008). That is

$$Y = kL\Lambda_{\mathrm{B}} \sum r^2 \tag{17.2}$$

where k is the empirical erodibility value, L is the total slope length (m), Λ_{B} is the dimensionless local slope gradient indicated by an evaluation at the slope base, taken over the frequency distribution of daily rainfall storm events in an average year, and r is an estimate of the accumulated runoff.

A total of 93 climatic, land use/cover, topographic, and physiographic parameters were assembled and used to simulate the 2002/2003 season cycle on a monthly time scale with all parameters prepared at, or re-sampled to, 90 m resolution (see Munthali et al. 2011).

Basin storage of sediment is an important variable in sedimentation management practices, as not all the sediment eroded from uplands is immediately delivered downstream (Nagle et al. 1999). Drainage density, which Bozali et al. (2008) define as the capability of a stream network to discharge total rainfall, and hence sediment, from its watershed, is one indicator of basin storage, and is quantified as

$$D_s = \frac{L_s}{A} \tag{17.3}$$

where D_s is the drainage density, L_s is the total stream length of the system, and A is the total area of the watershed.

A multi-criteria evaluation was then carried out to identify the critical areas requiring prioritized conservation management based on the following factors: (1) distance to the river channel; (2) slope; (3) the estimated PESERA outputs of runoff as a delivery medium (Schumm 2009) and sediment load (Fig. 17.1).

17.3.3 Results and Discussion

17.3.3.1 Sediment Generation

The estimated mean annual sediment generation varied considerably in magnitude, but showed no significant spatial variability (Fig. 17.2). Worrying levels are confined to the upper catchment, with most of the areas having an estimated sediment potential of over *25 tonnes per hectare per year* (t ha^{-1} $year^{-1}$) (Munthali et al. 2011). With high topographic percentage rises in the upper catchment and extreme magnitude variability of the rainfall both in space and time, infrequent but heavy rains tend to be responsible for moving much of the sediment generated in the tropical catchment (Munthali et al. 2011).

The monthly distribution of sediment deliverable to the reservoir follows the rainy season (see Fig. 17.3). No vegetative cover loss (0%) would have been an ideal control, but that would not have been realistic, and the model opted for 5% cover. Although the model has under-estimated the sediment yield (see Fig. 17.3), the results compare very well considering that the observed sediment yield translation used the observed monthly maximum discharge values (Munthali et al. 2011). The actual observed sediment is expected to be less than depicted, thereby coming closer to the model estimates. Despite the fact that some observed sediment concentration data were missing for the dry months (June to November), the available data sufficed in this analysis as it was for the critical window period—the rainy season. The model estimated monthly series translate into an average annual estimated sediment input of 9,000,000 t $year^{-1}$, which is equivalent to 3.4 Mm^3 $year^{-1}$, and which raises concerns about sedimentation (Munthali et al. 2011). These results point to a high sedimentation risk, and compare well with relative sedimentation estimates of similar tropical catchments in India and the Philippines (Nagle et al. 1999).

17.3.3.2 Soil Erosion Management

Sustainable watershed rehabilitation and management do not only depend on the quantified magnitudes of sediment generated, but also on an awareness of the sediment residence time in the catchment (Nagle et al. 1999). The drainage density

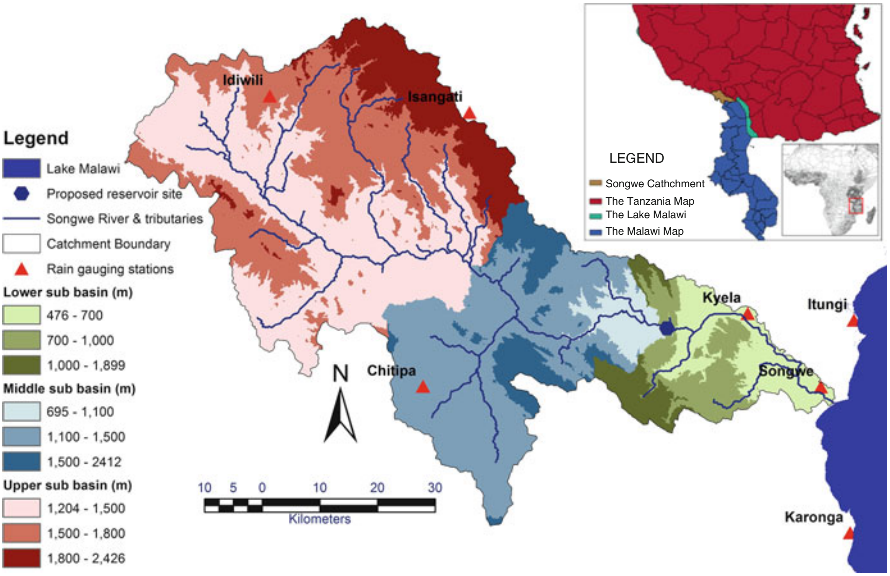


Fig. 17.1 Physiographic sub-basins showing elevation distribution above sea level and the proposed reservoir site (*Source*: Munthali et al. 2011)

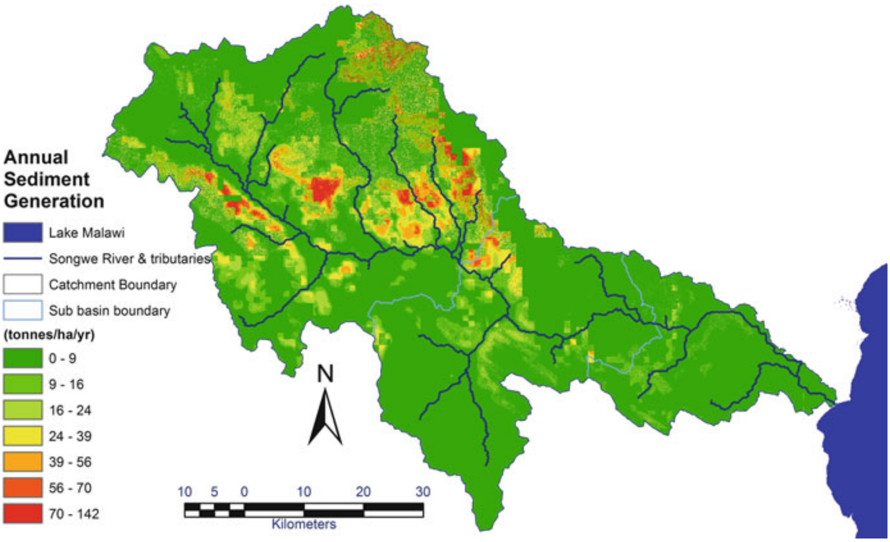


Fig. 17.2 Estimated average annual sediment generation (t ha^{-1} year^{-1}) (*Source*: Munthali et al. 2011)

was calculated to be very low at 0.1126, signifying that the available sediment in the watershed has a long residence time (Munthali et al. 2011). This makes the sediment generated in the catchment available for transport even after further erosion has been prevented upstream.

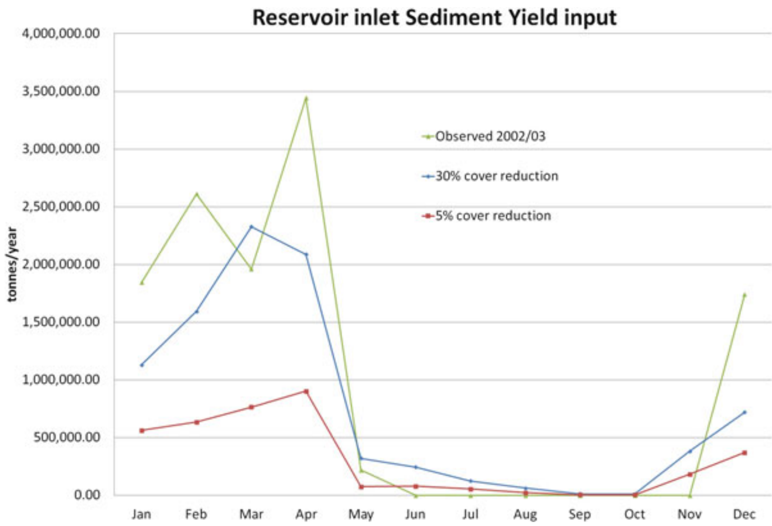


Fig. 17.3 Estimated sediment yields delivered to the proposed reservoir site, and observed sediment concentration at the Mwandenga Gauging Station (*Source*: Munthali et al. 2011)

Furthermore, a determination of whether sediments are from human or geological origin is of particular importance to reservoir capacity and storage sustainability (Nagle et al. 1999). This determines the type of sediment expected, e.g., large from landslides and/or fine from agricultural and built-up land, and permits a focused catchment conservation strategy.

Dams are built with the provision of a portion of the reservoir to hold the incoming sediment loads which are projected to occur during the economic life of the dam; referred to as "dead storage" (Haregeweyn et al. 2006; Robert 1973). This storage is in the deepest portion of the reservoir, usually close to the dam, between the level of the water intake and the bed of the stream. "Live storage," on the other hand, refers to the stored water that will actually be used (Nagle et al. 1999). While fine sediment is carried as a suspended load further into the reservoir and settles into the planned dead storage area, coarse sediment is carried as bed load and is deposited at the reservoir inlet, thus taking up part of the live storage area (Dunne 1988).

By overlaying the result in Fig. 17.2 with the land use/cover map for the study area, it was found that a large proportion of the sediment originates from built-up village land i.e., rural village settlements, footpaths, and all open and non-arable areas (see Fig. 17.4). Dunne (1979) observed that although there have been very few studies of paths and village settlements in estimates of the actual sources of sediments, these features generate a large proportion of the sediment yield of catchment areas. An increase in the population density in the study area (Chimphamba et al. 2006) has made the network of roads and paths denser and the area covered by the settlements larger. In the upper and middle sub-basins, the

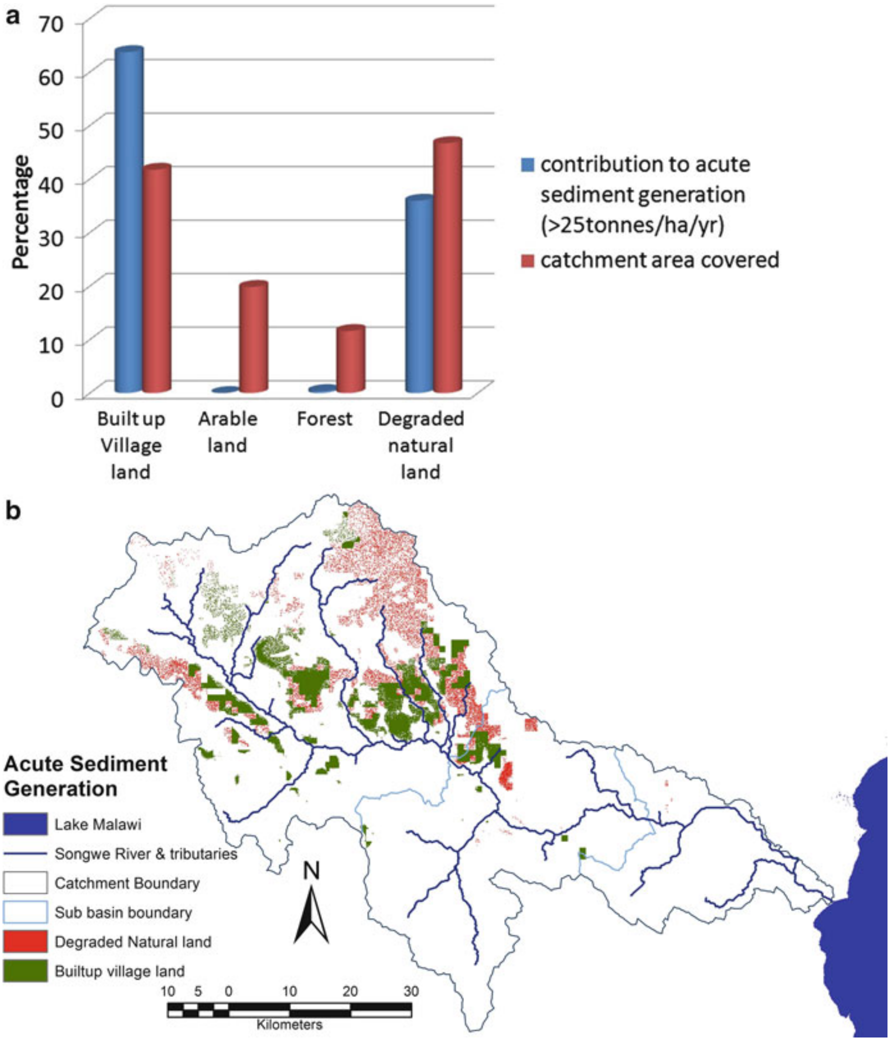


Fig. 17.4 Acute sediment generation per land use. (**a**) Proportion of each land-use type (*Source*: Munthali et al. 2011). (**b**) Spatial distribution for the highest sediment generating types

networks traverse steeper gradients of the terrain. Whether fine and/or coarse sediments are generated (which depends on the prevailing catchment conditions), the combined contribution from built-up village land was significantly over half of the total found in the study area. This is similar to the results from tropical watersheds elsewhere (Nagle et al. 1999).

It has been reported that intensive cut-and-burn agriculture (Chimphamba et al. 2006) has left many natural environments degraded. The tropical storm rainfall conditions subject the degraded natural land to mass wasting, which ends up constituting a large proportion of the bed load sediment of the stream channel (Munthali et al. 2011).

17.4 Discussion and Conclusion

Watersheds are life lines to large populations, and are critically so to rural communities in the developing world. However, most watershed catchments continue to degrade,with the majority of them producing large amounts of sediment resulting from uncontrolled soil erosion (Munthali et al. 2011; Nagle et al. 1999). It has been established that unstable terrain and climatic features produce erosion rates that are so high that human activities are ineffective in controlling the resultant sediment yields (Mahmood 1987). This begs the question of whether the sediments are of human or geological origin (Nagle et al. 1999). It has been found that despite the fact that reducing soil erosion from agricultural land is important for many reasons, focusing on this has been known to have a limited impact on reducing sedimentation and improving overall watershed management (Munthali et al. 2011; Nagle et al. 1999). This is a particular problem in cases where built-up village land and degraded natural land in tropical mountainous areas result in critically high soil erosion areas that pose a two-tier watershed sustainability problem: (1) there is a direct potential for the "dead" storage area to be quickly filled from the huge amount of fine sediment generated in the case of built-up village land; (2) there is a shift in sedimentation to "live" rather than "dead" storage due to the coarser sediment generated from the degraded natural land (Munthali et al. 2011). The latter is a case of mass wasting, which Ahmad et al. 1993 estimated to be around 81% of the total estimated sediment transport in the forested basins of Puerto Rico. Similar huge sedimentation observations have been made in regions where little or no human habitation exists (Bruijneel 1990; Pearce 1986).

Nevertheless a process model, and particularly a coarse-scale model such as PESERA, has a number of inherent disadvantages compared with simpler models (Kirkby et al. 2008). The worst of these is their huge demand for input data (Munthali et al. 2011). Technically, these process-based models face an inevitable concentration on the relevant dominant processes that are most widespread, in this case infiltration excess over-land flow, so that erosion by saturation over-land flow, for example, is less well estimated (Kirkby et al. 2008).

However, although fraught with limitations even in the best of circumstances, the geospatial identification of sources of soil erosion and sediment serves to avoid guesswork in planning management programs for watersheds (Nagle et al. 1999). It also develops an awareness of whether the sources are amenable, and whether the sedimentation control approaches can produce quantitatively verifiable results when the critical sources identified are prioritized (Munthali et al. 2011). This systematic geospatial development and analysis of sedimentation serves to improve both the planning of watershed development schemes and the allocation of resources towards reducing sedimentation. This is becoming important, especially in the developing world where finances and technical resources are limited, while erosion control methods need to be directed to areas where they would be most effective (Nagle et al. 1999). However, computer-based geospatial analysis must be coupled with accurate groundwork on the sources of sediment and erosion in order to guarantee the suitability and success of the evaluation (Munthali et al. 2011).

References

Ahmad R, Scatena FN, Gupta A (1993) Morphology and sedimentation in Caribbean montane streams: Examples from Jamaica and Puerto Rico. Sedimentary Geomorphology 85:157–169

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of catchment hydrology. Hydrol Sci J 24:43–69

Bozali N, Yüksel A, Akay AE (2008) Determining the main factors effecting the sediment yield from Derindere watershed of Sir Dam in Kahramanmaras by using GIS techniques. Int J Nat Eng Sci 2:29–32

Bruijneel LA (1990) Hydrology of moist tropical forests and effects of conversion: a state of knowledge review. International Association of Hydrological Sciences, Paris, p 224

Chimphamba JB, Mwakalila S, Kanyanga J (2006) Assessment of Cut-and-Burn farming system on the management of Songwe River: a trans-boundary watershed between Tanzania, Zambia and Malawi. Presented at United Nations/Zambia/European Space Agency Regional Workshop on the Applications of Global Navigation Satellite System Technologies for Sub-Saharan Africa, Lusaka, Zambia. Retrieved on 1 June 2009, from http://www.unoosa.org/pdf/sap/2006/zambia/presentations/04-01-03.pdf

Commission of the European Communities (2006). Proposal for a Directive of the European Parliament and of the Council Establishing a Framework for the Protection of Soil and Amending, Directive 2004/35/EC

Darboux F, Gascuel-Odoux C, Davy P (2002) Effects of surface water storage by soil roughness on overland-flow generation. Earth Surf Process Landforms 27:223–233

de Roo APJ (1996) LISEM: a single-event physically based hydrological and soil erosion model for drainage basins. 1. Theory, input and output. Hydrol Process 10:1107

Dunne T (1979) Sediment yield and land use in tropical catchments. J Hydrol 42:281–300

Dunne T (1988) Geomorphologic contributions to flood control planning. In: Baker VR, Kochel RC, Patton PC (eds) Flood geomorphology. Wiley, New York, pp 421–438

Elwell HA (1981) A soil loss estimation technique for southern Africa SLEMSA. In: Morgan (ed) Soil conservation problems and prospects. Wiley, Chichester, pp 281–292

Fried JS, Brown DG, Zweifler MO, Gold MA (2000) Mapping contributing areas of stormwater discharge to streams using terrain analysis. In: Wilson JP, Gallant JC (eds) Terrain analysis, principles and applications, 1st edn. Wiley, New York, pp 183–203

Friedkin JF (1945) A laboratory study of the meandering of alluvial rivers. US Waterways Experimental Station, Vicksburg

Gobin A, Jones RJA, Kirkby MJ, Campling P, Govers G, Kosmas C (2004) Indicators for pan-European assessment and monitoring of soil erosion by water. Environ Sci Pol 7:25–38

Haboudane D, Bonn F, Royer A, Sommer S, Mehl W (2002) Land degradation and erosion risk mapping by fusion of spectrally based information and digital geomorphometric attributes. Int J Rem Sens 23:3795–3820

Haregeweyn N, Poesen J, Nyssen J, De Wit J, Haile M, Govers G, Deckers S (2006) Reservoirs in Tigray (Northern Ethiopia): characteristics and sediment deposition problems. Land Degrad Dev 17:211–230

King D, Jones RJA, Thomasson AJ (eds) (1995) European land information systems for agro-environmental monitoring. Office for Official Publications of the European Communities, Luxembourg, EUR 16232 EN

Kirkby MJ, Irvine BJ, Jones RJA, Govers G (2008) The PESERA coarse scale erosion model for Europe. I. Model rationale and implementation. Eur J Soil Sci 59:1293–1306

Le Bissonnais Y, Cerdan O, Lecomte V, Benkhadra H, Souchere V, Martin P (2005) Variability of soil surface characteristics influencing runoff and inter-rill erosion. Catena 62:111–124

Magrath WB, Doolette JB (1990). Strategic issues in watershed development. In: Doolette JB, Magrath WB (eds) Watershed development in Asia: strategies and technologies. Technical Paper Number 127. World Bank, Washington, pp 1–25

Mahmood K (1987) Reservoir sedimentation: impact, extent, mitigation. Technical Paper Number 71. World Bank, Washington, p 118
Morgan RPC, Quinton JN, Rickson RJ (1994) Modelling methodology for soil-erosion assessment and soil conservation design – the Eurosem approach. Outlook Agr 23:5–9
Mulder T, Syvitski JPM (1996) Climatic and morphologic relationships of rivers. Implications of sea level fluctuations on river loads. J Geol 104:509–523
Munthali KG, Irvine BJ, Murayama Y (2011) Reservoir sedimentation and flood control: using a geographical information system to estimate sediment yield of the Songwe River Watershed in Malawi. Sustainability 3(1):254–269
Nagle GN, Fahey TJ, Lassoie JP (1999) Management of sedimentation in tropical watersheds. Environ Manage 23:441–452
Nearing MA, Foster GR, Lane LJ, Finkner SC (1989) A process-based soil-erosion model for USDA-water erosion prediction project technology. Trans Am Soc Agric Eng 32:1587–1593
Norplan A.S. in Association with COWI, DHI Water & Environment, WandPES (2003). NORPLAN: a Joint Project between the Government of the Republic of Malawi and the Government of the United Republic of Tanzania. Final Feasibility Study Report; Preliminary Environmental Impact Assessment Main Report, vol 4A
Pandey A, Chowdary VM, Mal BC, Billib M (2008) Runoff and sediment yield modeling from a small agricultural watershed in India using the WEPP model. J Hydrol 348:305–319
Pearce AJ (1986) Erosion and sedimentation. Prepared for the workshop on ecological principals for watershed management, 9–11 April 1986, East-West Center, Honolulu, p 18
Pilotti M, Bacchi B (1997) Distributed evaluation of the contribution of soil erosion to the sediment yield from a watershed. Earth Surf Process Landforms 22:1239–1251
Renard KG, Foster GR, Weesies GA, Porter JP (1991) RUSLE. Revised universal soil loss equation. J Soil Water Conserv 46:30–33
Rober SI (1973) Sedimentation. In: Design of Small Dams, United States Department of the Interior Bureau of Reclamation. Water Resource Technical Publication, Oxford and IBH, New Delhi
Roose E (1996) Land husbandry – components and strategy. Ecuador: 70 FAO Soils Bulletin. Retrieved on 19 September 2011, from http://www.fao.org/docrep/T1765E/T1765E00.htm
Schumm SA (2009) Streamflow and sediment yield; Chicago, IL, USA: Encyclopedia Britannica. Retrieved on 20 July 2009, from http://www.britannica.com/EBchecked/topic/504801/river/29099/Factors-that-influence-sediment-yield
Smith RE, Goodrich DC, Quinton JN (1995) Dynamic, distributed simulation of watershed erosion – the Kineros2 and Eurosem models. J Soil Water Conserv 50:517–520
Tamene L, Park SJ, Dikau R, Vlek PLG (2006) Reservoir siltation in the semi-arid highlands of Northern Ethiopia: sediment yield-catchment area relationship and a semi-quantitative approach for predicting sediment yield. Earth Surf Process Landforms 31:1364–1383
U.S. Department of Agriculture-Soil Conservation Service (1994) Summary Report: 1992 National Resources Inventory. Statistical Laboratory, Iowa State University, Ames
Wellmeyer JL, Slattery MC, Phillips JD (2005) Quantifying downstream impacts of impoundment on flow regime and channel platform, lower Trinity River, Texas. Geomorphology 69:1–13
WWF (World Wide Fund) (2009) Flood preparedness study for the Lower Songwe Basin, Main Report, vol I
Williams JR (1982) The physical components of the EPIC model. In: El-Swaïfy, Moldenhauer & Co (eds) Soil erosion and conservation. ISCO 3. SWC. Sol of America, Ankeny pp 273–284
Wischmeier WH, Smith DD (1958) Rainfall energy and its relationship to soil loss. Trans Am Geophys Union 39:285–291
Wischmeier H, Smith DD (1978) Predicting rainfall erosion losses – a guide for conservation planning. Agriculture Handbook 537. U.S. Department of Agriculture, Washington

Index

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7, © Springer Japan 2012